(1)

COPING WITH NOVELTY AND HUMAN INTELLIGENCE: THE ROLE OF COUNTERFACTUAL REASONING

ONR Contract N00014-85-K-0589 - FINAL REPORT

Robert J. Sternberg
Yale University
New Haven, CT

COPING WITH NOVELTY AND HUMAN INTELLIGENCE:
THE ROLE OF COUNTERFACTUAL REASONING

ONR Contract N0001485K0589 - FINAL REPORT

Robert J. Sternberg
Yale University

**Coping with Novelty and Human Intelligence:
The Role of Counterfactual Reasoning**

ONR Contract N0001485K0589 Final Report

Robert J. Sternberg
Yale University

This Final Report is divided into four main sections. In the first section, I describe in some detail a subset of the empirical research I have done with my collaborators under the contract. Rather than attempt to describe every study, I have described selected studies in detail. These studies show some of the range of work we have accomplished during the contract. In the second section, I present a list of the publications that have ensued from 1985 to 1988, as well as publications that are in press. These are the years during which the contract was held. In the third section, I list presentations I have given during the years of the contract. And in the fourth and final section, I give the final budgetary accounting for the contract. Keywords: Intelligence (Humans) - Hypotheses, Psychomotor tests, Learning, Aptitudes test, Cognition, Information processing, experimental psychology (Sdw)

Section 1

Coping with Novelty in Human Intelligence: An Empirical Investigation[1]

Kuhn (1970) has observed that the truly influential scientists are those who see things differently from the way others see things: They are the scientists who create new paradigms rather than following existing ones. In other words, they can take an old problem and see it in a way that is different from the way other scientists see it. For example, Chomsky (1965) created a new paradigm in linguistics when he introduced the concept of transformational grammar, seeing language, a familiar entity, in a way no one had seen it before.

Kuhn's observation applies not only to science. In the investment world, those investors who are considered to be unusually successful by virtue of the outstanding performance of their investment portfolio are almost always, to greater or lesser degree, "contrarian" investors (Dreman, 1982). The reason for this phenomenon is that a good investment generally stops being good just as soon as everyone recognizes that it is good. In the stock market, for example, the only way to make money on a stock is to recognize that it is undervalued before everyone else recognizes it to be undervalued. As soon as everyone else recognizes that fact, they buy the stock and it is no longer undervalued, as its price has now risen.

The observations of scholars about science and the investment world apply in everyday life as well. When questioning someone in order to establish her intelligence or general competence, we often throw the person a "curve ball," meaning that we ask her a question that is different in kind from the question that is typically asked of people in a given situation. If she is able to answer ordinary questions, but not ones that require some degree of flexible thinking or insight, we may conclude that she is only somewhat intelligence or competent, or even that she is pedestrian. Indeed, in their studies of people's conceptions of intelligence, Sternberg, Conway, Ketron, and Bernstein (1981) found that people valued as intelligent the ability to see problems in a new and different way.

Psychologists as well as laypersons have recognized for some time the importance of coping with novel kinds of tasks or situations to intelligence for a long time. For example, psychometric tests of intelligence often include items that measure a person's ability to cope with novelty, and the theory of intelligence proposed by Cattell and Horn (Cattell, 1971; Horn and Cattell, 1966) distinguishes between two kinds of intelligence—crystallized and fluid. Crystallized intelligence is the set of products representing accumulations of the processes of fluid intelligence, which are the processes used in thinking in new and different ways. Tests measuring fluid ability would be abstract analogies, classifications, series, or matrices, all of which require one to cope with novel kinds of problems. Raaheim (1974) viewed the ability to cope with relative novelty as the key ingredient of intelligence, a view taken as well by Sternberg (1981, 1982, 1985), whose triarchic theory of intelligence holds that the ability to cope with relative novelty is one of three key ingredients to human intelligence (see Sternberg, 1985). Piaget (1972), too in a different tradition, measured intelligence by presenting children with novel kinds of tasks, such as the balance-beam task or any of a number of conservation tasks.

Although laypersons and psychologists alike have recognized the importance of coping with novelty to intelligence, oddly enough, there appears to be no direct experimental test of the hypothesis that a relatively novel version of a given task will provide significantly better assessment of a person's intelligence than will a non-novel version of that task. Although a number of investigators, including ourselves (Sternberg & Gastel, in press), have shown the relevance of the ability to cope with novelty to intellectual performance (see also Davidson & Sternberg, 1984; Sternberg, 1982; Sternberg & Davidson, 1982; Tetewsky & Sternberg, 1986), the line of argument has been that tasks that require coping with novelty correlate fairly highly with psychometric tests of fluid intelligence. None of these studies have shown significantly higher correlations for the tasks requiring coping with novelty than for tasks that do not require this ability.

The purpose of this investigation was directly to test the hypothesis that the ability to cope with a relatively novel version of a task is more highly related to fluid intelligence than is the ability to cope with a non-novel version of that task. In other words, this study seeks directly to test the hypothesis that the ability to cope with novelty is a key ingredient of intelligence. We tested this hypothesis by presenting subjects with a verification task that requires subjects to make either a familiar presupposition or a novel one. We hypothesized that requiring the subject to reason with a novel presupposition would increase the extent to which a task measures fluid intelligence.

<u>Subjects</u>

Subjects in the main study were 50 Yale undergraduates (17 females, 33 males), all of whom received course credit for their participation in the experiment, which took roughly two hours. In addition, 60 summer students at Yale provided ratings in a prestudy that were used in order to validate the materials for the main study.

<u>Materials</u>

There were two basic kinds of materials: a verification task and psychometric ability tests. The verification tasks comprised the main experimental stimuli, whereas the ability tests were used to validate the main hypothesis of the study regarding the relationship of coping with novelty with fluid intelligence.

<u>The verification task</u>. The verification task consisted of 232 statements, each of which a subject had to identify, as rapidly as possible, as being either true or false. Statements were grouped into sets of from five to eight individual statements, and each of these sets was preceded by a presupposition, which subjects were instructed to assume was true. For half of 34 different sets of items, the presupposition was one that was true in the everyday world; for the other half of the item sets, the presupposition was counterfactual. A given set of items was paired with both a familiar and a counterfactual presupposition. Half of the subjects saw a given set of items with the familiar presupposition, and half saw that set with the counterfactual presupposition. Table 1 shows some representative sets of items from the experiment, paired with both their familiar and their counterfactual presupposition. Of course, a given subject saw each of the items with only one of the two possible presuppositions and did not see the keyed answer.

Table 1 illustrates that for statements with familiar presuppositions, the correct answer to the verification task is the same as it would be in the ordinary, everyday world. However, for item sets with a counterfactual presupposition, the keyed answer was the same as that in the everyday world for only half the total number of items. For this half of the items with the counterfactual presupposition, the presupposition was irrelevant to verification of the statement. For the other half of the items with a counterfactual presupposition, however, the presupposition was relevant, and hence changed the keyed response. Note, then, that the task version with the counterfactual presupposition required a certain kind of flexibility that the task version with the familiar presupposition did not require, as the former version of the task required a different answer from the normal, everyday one for half the items, whereas the version of the task with familiar presuppositions never required an answer that was different from the normal, everyday one.

Insert Table 1 about here

The subjects in the prestudy were used to verify that our keying of responses was the same as that which would be rendered by the subjects actually performing the task. The prestudy subjects were asked to assume that each presupposition was true, and then to mark each following statement as true or false, based on the presupposition. Subjects were also asked to provide a confidence rating for each statement indicating how sure they felt they were of the correctness of their answer (on a 1=low, 7=high scale). Questionnaires consisted of 36 sets of items, half with familiar presuppositions and half with counterfactual presuppositions. Two different forms of the questionnaire were administered to two groups of subjects. The questionnaires differed in that items having counterfactual presuppositions in the first form appeared with factual presuppositions in the second form, and vice versa. Although the presuppositions differed between forms, the statements following them were identical. Items were retained for the final study only if there was clear agreement regarding what should be the keyed answer.

Whereas in the prestudy the items were administered with no time pressure, subjects in the main study were aware that they would be timed in performing the tasks, and were told to be as accurate as possible without making errors.

Psychometric tests. All subjects in the final study received four psychometric ability tests in a fixed order: (a) the Letter Sets Test from the French Kit of Reference Tests for Cognitive Factors (Ekstrom, French, & Harman, 1976), (b) the Syllogisms Test from the California Test of Mental Maturity, Level 5 (Sullivan, Clark, & Tiegs, 1963), (c) the Cattell Culture-Fair Test of g, Scale 3, Form A (Cattell & Cattell, 1963), and (d) the Finding A's Test from the French Kit of Reference Tests for Cognitive Factors (Ekstrom, French, & Harman, 1976). The first three of these ability tests are well-validated measures of fluid intelligence included for purposes of convergent validation of the experimental hypothesis. The last test was a test of perceptual-motor speed, included for purposes of discriminant validation.

Design

The main dependent variable was response time to each of the statements to be verified. A secondary dependent variable was error rate. Psychometric ability-test scores were used as dependent variables for purposes of convergent and discriminant validation. The main independent variables were presupposition type (novel, non-novel) and type of response (true, false).

Procedure

In the prestudy, subjects received sets of statements paired with either familiar or counterfactual presuppositions, and had to indicate for each statement whether it was true or false, given the presupposition, and also what their confidence rating was (on a 1-7 scale) for their labeling of the statement as true or false. Item sets were presented in random order. In the main experiment, subjects also received items sets in random order, followed at the end of the experiment by the psychometric ability tests in fixed order (in a separate group session). At the beginning of the experimental session, each subject was orally briefed on the general content of the experiment, and then presented with a consent form to sign. Stimuli were presented on a MacIntosh personal computer. First, instructions appeared on the computer screen. Then the subjects saw four practice items. On these practice items, the computer informed the subject whether each response was correct or incorrect. However, no feedback was provided for the rest of the items. Half way through the computer session, subjects were given an opportunity for a break, after which they continued with the second part of the experimental items. Ability tests, which were all in paper-and-pencil format, were administered after all verification statements were completed.

Before each subject began, all item sets were randomized and all statements within each item set were randomized in order. Each subject saw all 34 item sets, including 232 statements. Odd-numbered subjects saw items from a first batch with counterfactual presuppositions and items from a second batch with familiar presuppositions; even-numbered subjects saw the reverse pairing.

Each item appeared as follows: First the presupposition was presented, prefaced by the phrase "Assume that...." The subject had as long as he or she desired to read the presupposition. When ready, the subject would press the space bar, which would initiate the presentation of a statement to be verified. The subject responded to the statement by pressing a key on the computer board. The subject's response and response time were recorded. After a half-second delay, the next statement would appear, and so on. All statements appeared in bold face type, with the presupposition written above in normal type in parentheses. The presupposition appeared for each statement so the subjects would not have to memorize the presuppositions.

Results

Table 2 presents the means for each of the main conditions, and Table 3 presents the results of the analyses of variance. As would be expected, novel items were more difficult than familiar items, as indexed both by higher response times and higher error rates. False items were more difficult than true ones. The correlation of response times with error rates across items was .26 ($p < .001$).

Insert Tables 2, 3, and 4 about here

Table 4 shows correlations of response times and error rates with psychometric ability test scores. The task shows good convergent and discriminant validity, suggesting that it is a good measure of fluid intelligence but that it does not measure perceptual-motor speed.

The usefulness of the verification task for measuring fluid abilities is shown by a stepwise multiple regression of the psychometric ability tests on the experimental task scores. For both novel and non-novel items, two of the three ability tests enter into the multiple regression: Letter Sets and Syllogisms. For novel response times, the multiple correlation was .73 with the two ability tests, $p < .001$. For the non-novel items, the multiple correlation was .68, $p < .001$. Thus, the verification task appears to tap well the same construct as is tapped by conventional fluid-ability tests.

The critical correlations for the experiment are those that test the hypothesis that the novel condition provided better assessment of fluid intelligence but not perceptual-motor speed than did the non-novel condition. In order to test this hypothesis, a difference score was computed for each subject whereby the subject's average response time for non-novel was subtracted from his or her average response time for novel items. This difference score reflects the increment in time to respond for novel items (i.e., items with counterfactual presuppositions) over non-novel items (i.e., items with familiar presuppositions). The correlation of this difference score with the Letter Sets Test was -.34, $p < .02$. The correlation of the difference score with the Syllogisms Test was -.38, $p < .01$. The correlation with the Cattell Test was -.15, $p > .05$. Finally, the correlation with the Crossing out A's Test was .04, $p > .05$. Thus, all three correlations of the difference score with the psychometric ability tests were negative, as would be predicted by the main hypothesis of the experiment. Two out of the three correlations were statistically significant. The correlation of the perceptual-motor speed test with the difference score was trivially positive. In other words, the main hypothesis of the study was generally confirmed.

Discussion

The goal of this experiment was directly to test the hypothesis that coping with novelty is an important aspect of intelligence. The test of the hypothesis was direct in the sense that subjects received both a relatively novel and a relatively non-novel version of a sentence verification task, and it was shown that the difference score comparing response times for the two tasks was significantly correlated with psychometric measures of fluid abilities. In other words, the increment in time associated with the novel versus non-novel version of the task significantly predicted fluid ability. We believe that this result constitutes the most direct test to date of the hypothesis that coping with novelty is an important aspect of fluid intelligence, although it certainly does not constitute the only test. Moreover, we recognize that no single task can provide a broad test of the hypothesis, as there are many kinds of novelty, and the kind of novelty involved in the verification task with counterfactual presuppositions is only one of many possible kinds.

Our intuition that the ability to see things in a new and different way is an important ingredient of intelligence is confirmed by this study. This is not to say, however, that all intelligent people have the ability to cope well with novelty. To the contrary, some people may be very intelligent just so long as their everyday presuppositions can be taken for granted. The intelligence with which they perform, however, may breakdown when novelty is introduced. Conversely, some people may be at their best in coping with novelty, and look less intelligent in comparison with others when confronted with more mundane tasks (Sternberg, 1985). In understanding intelligence, then, we need to view it in its diversity, and to recognize that intelligence has many aspects, an important one of which is the ability to cope with relative novelty.

If Dancers Ate Their Shoes: Inductive Reasoning with Factual and Counterfactual Premises[2]

In our everyday lives, we must frequently make inductive judgments. Most of these judgments are routine, requiring standard reasoning from familiar premises. But occasionally we must assume that something is true about which we may have little or no information, or which may even be counterfactual. Suppose, for example, we view Nicaragua as having a socialist government interested only in self-determination. Does the standard U.S. government analogy between Nicaragua and Cuba still hold? Or suppose we believe that a leading clinical psychologist has three maladjusted children. Can this psychologist's clinical advice still be trusted?

An example of the need for such potentially counterfactual inductive reasoning occasionally arises in detective work, as shown in Isaac Asimov's (1979) short story, "The Singing Bell." The question faced by the world-famous professor of extraterrestrial phenomena, Dr. Urth, is whether Louis Peyton was on the moon recently, where he allegedly murdered Albert Cornwell. Solving the problem requires Dr. Urth to reason about the effects of unfamiliar gravitational levels.

The motive for the murder would have been to obtain a large quantity of a rare and valuable life form called the Singing Bell. Peyton denies both the murder and having been on the moon. Dr. Urth asks the accused, Peyton, whether he respects Singing Bells. Peyton replies that he does--too much to break one. At this point, Peyton gently strokes the Singing Bell he is holding. Dr. Urth suddenly commands Peyton to toss the Singing Bell to him. Peyton automatically tosses the Bell. It travels a short arc one-third the way to Urth, curves downward, and shatters on the floor. Dr. Urth then concludes:

> Surely the matter is now obvious. The fact that Mr. Peyton could so egregiously misjudge the toss of an object he obviously valued so highly could mean only that his muscles are not yet readjusted to the pull of Earthly gravity. It is my professional opinion, Mr. Davenport, that your prisoner has, in the last few days, been away from Earth. He has either been in space or on some planetary object considerably smaller in size than the Earth--as, for example, the Moon. (p. 237)

In the above story, Dr. Urth thinks in terms of what would follow if an unfamiliar situation were true. Such reasoning with unfamiliar information can be readily incorporated into inductive reasoning problems. For example, assume that dinosaurs are kinds of fruit juices. Given this premise, what word would come next in the following series: SUBSTANCE, LIQUID, DRINK, FRUIT JUICE, (a) BRONTOSAURUS, (b) COFFEE, (c) VEGETABLE, (d) GLASS?

Our attempt to explore reasoning with different kinds of information is part of an ongoing program of research in which we explore the relations between task novelty and information processing in tasks requiring human intelligence. This program of research is motivated by the notion that the ability to deal with relative task novelty is a particularly crucial aspect of human intelligence (Sternberg, 1981, 1985; see also Raaheim, 1974). In other words, tasks that present subjects with intermediate amounts of novelty may be good measures of intelligence.

This assertion is part of the triarchic theory of intelligence (Sternberg, 1985). According to this theory, intelligence comprises components of information processing that people employ in deciding how to adapt to, shape, or select their environment. Particularly relevant to the present experiment is the experiential subtheory, which asserts that intelligence is measured to the extent that a particular task is relatively novel (as in the present experiment) or requires automatization of information processing. In this experiment we seek to test the novelty aspect of this subtheory through the use of counterfactual novelty in inductive reasoning problems.

In this experiment we also explore the ability to sift relevant from irrelevant information. This ability, which we call selective encoding, is one of three abilities (the others are selective combination and selective comparison) which we have proposed as basic to insight (Sternberg & Davidson, 1982, 1983). This theory of insight is also part of the experiential subtheory of the triarchic theory of intelligence.

The scheme of the program of research is shown in Table 5. Our

---

Insert Table 5 about here

---

research on induction has proceeded through four overlapping phases. In the first phase, we and others attempted task decomposition to understand the information-processing components underlying performance on familiar, IQ-test-like inductive-reasoning items (Sternberg, 1977; Sternberg & Gardner, 1983; see also Mulholland, Pellegrino, & Glaser, 1980). Performance on tasks such as analogies, classifications, and series completions was decomposed into its elementary components of processing, enabling us to ascertain the latencies of such components, and the strategies into which they were combined. In a second phase, we investigated tasks that were, in some sense, at the opposite extreme, namely, unfamiliar and un-IQ-test-like. In one kind of problem, the conceptual projection problem, subjects were required to predict the future state of an object, given incomplete information about its present state. Objects in these experiments were other-worldly. For example, the people of the planet Kyron could be either born young and die young, born old and die old, born young and die old, or born old and die young (Sternberg, 1982; see also Tetewsky & Sternberg, 1986). In a second kind of problem, the insight problem, subjects had to solve what appeared to be standard mathematics problems, but what were in fact insight problems that could not be solved by routine mathematical formulae (Davidson & Sternberg, 1984; Sternberg & Davidson, 1982). In a third phase, we investigated problems that were familiar but un-IQ-test-like--inductive predictions into the future (when will a bottle of milk spoil?) and postdictions into the past (when did the bottle of milk spoil?) (Kalmar & Sternberg, 1985). The present experiment represents work in a fourth phase of research.

In the fourth phase, we have been investigating problems that are IQ-test-like, but unfamiliar. Such problems have the theoretical advantage that they may measure the ability to cope with relative novelty, which seems quite important to intelligence (Sternberg, 1985), and may also be practical for use both in laboratory experiments and on tests of human intelligence, broadly defined. These problems involve a mix of items, some requiring reasoning based on facts and others requiring reasoning based on counterfactual ("novel") premises. In a first experiment in this series, Marr and Sternberg (1986) found that both gifted and nongifted students in grades 6, 7, and 8 gave significantly more attention to novel information than to familiar information in the test problems, which were analogies preceded by a cue that could be either novel (e.g., sparrows play hopscotch) or familiar (e.g., pistols are weapons), and either relevant or irrelevant to analogy solution. Gifted students, however, gave significantly less attention to irrelevant novel information than did nongifted students, but did not differ from the nongifted students in their attention to relevant novel information.

The Marr and Sternberg (1986) experiment addressed differences in outcomes of information processing of factual and counterfactual analogies for gifted and nongifted children, but did not examine the question of just what information processing is used, nor did it investigate information processing in other types of problems. Moreover, the work was limited to grade-school children. In the present experiment, we seek to confront head-on the question of how adults process information in solving factual and counterfactual analogies, classifications, and series problems. These three kinds of test items were not chosen arbitrarily: They are the three inductive item types most often used to measure general intelligence, and particularly, so-called fluid intellectual abilities (cf. Cattell, 1971; Horn & Cattell, 1966). By investigating three related but distinct item types, we can investigate whether the results generalize or whether they are particular to each item type.

Method

Subjects

Subjects were 60 Yale undergraduates--17 males and 43 females--in an introductory psychology course. Subjects participated toward fulfillment of a course requirement.

Materials

Materials were of two basic types, experimental tasks and psychometric tests.

Experimental tasks. Three different experimental tasks were used: analogies, classifications, and series completions. In each task, half of the items were uncued and half were precued. Subjects saw each item in either precued or uncued form, but not both, with half the subjects seeing a given item in precued form and half in uncued form. Each task item was associated with a precue of one of four precue types--familiar relevant, familiar irrelevant, novel relevant, or novel irrelevant--which appeared only in the precued condition. Items were not the same across precue conditions. Examples of each of these types of items for each task are shown in Table 6.

Insert Table 6 about here

Each subject received a total of 216 inductive-reasoning items. Those items that appeared in uncued form were simply preceded by a blank field. Subjects had as long as they wanted to look at the blank field, and then pressed a button which caused the induction item to appear. Those items that appeared in precued form were preceded by a premise that could be either familiar or novel, and either relevant or irrelevant. Subjects were first presented with the precue, and were given as long as they wished to read it. They then pressed a button, which resulted in the disappearance of the precue and the immediate appearance of the induction item.

Items were equally divided among verbal analogies, classifications, and series completions, equally divided again between uncued and precued, and equally divided again (for the precued items) among familiar relevant, familiar irrelevant, novel relevant, and novel irrelevant precues, as follows: An item was classified as either familiar, if its precue was factual and well-known, or novel, if its precue was counterfactual. The item was also classified as either relevant, if its precue gave information helpful to item solution, or irrelevant, if its precue was unhelpful. Irrelevant precues, although unhelpful, did contain information associatively related to the item stem, so that the precue could not be immediately recognized as irrelevant.

Psychometric tests. Five psychometric tests were administered to each subject: the Verbal Reasoning subtest of the Differential Aptitudes Test (Form T) (Bennett, Seashore, & Wesman, 1973), the Cattell Culture Fair Test of g (Scale 3) (Cattell & Cattell, 1963), the insight problems used by Sternberg and Davidson (1982), the Crossing Out A's Subtest of the French Kit of Reference Tests for Cognitive Factors (French, Ekstrom, & Price, 1963), and the Vocabulary Subtest of the French Kit (V-3).

Design

Dependent variables were response times and error rates for the individual items. Independent variables, all within-subjects, were test type (analogies, classifications, series completions), cueing (uncued, precued), and, for precued items, precue type: familiarity (familiar, novel), and relevance (relevant, irrelevant). Each item had one cue associated with it, which was either familiar relevant, familiar irrelevant, novel relevant, or novel irrelevant.

Procedure

Subjects first filled out an informed consent form. Instructions were then administered orally. Subjects received the experimental tasks followed by the psychometric tests. The three experimental tasks (analogies, classifications, and series completions) were administered via an Apple IIe microcomputer, in counterbalanced order across subjects, and psychometric tests followed in a fixed order in a later session. Subjects were informed that on precued items, the precue should be accepted as true and should be applied, if appropriate, to the item that followed. Subjects were told to respond as quickly and as accurately as possible. Items within task were blocked by precueing versus no precueing, with each task consisting of four blocks of trials, with precued and uncued blocks alternating. Within the precued condition, the four kinds of cues were mixed so that a subject could not know on a given trial what kind of cue to expect. The first two blocks of each task were each preceded by four practice items. On uncued items, the subject initially saw a blank screen; on precued items, the subject initially saw the precue. When the subject pressed RETURN, the precue (if any) disappeared, and the test item appeared. Pressing of one of four response buttons resulted in termination of the trial. The computer recorded response times and responses, and then presented the next item.

Models of Information Processing

How do subjects solve the kinds of factual and counterfactual induction items presented in this experiment? Such a question can be addressed by considering alternative models of information processing. We should note that our goal in this article is not to provide a relatively complete model of information processing for beginning-to-end solution of analogies, classifications, and series completions: Such a task has been undertaken in an earlier article (Sternberg & Gardner, 1983; see also Sternberg, 1985). We have not included the necessary task manipulations to test the Sternberg-Gardner model on the present data. It is possible, of course, that the task manipulations we did include changed information processing so as to vitiate the applicability of our earlier model, although we have no reason to believe that this was the case. Our goal here is to discuss those aspects of information processing that are distinctive to the solution of induction items with familiar or novel precues that may be either relevant or irrelevant to problem solution.

Three basic kinds of models could account for information processing during solution of the induction problems: the null model, additive models, and interactive models. All of the proposed models are serial. We recognize, of course, that actual processing may not be strictly serial, and hence that our models may oversimplify the complex information processing subjects actually did. But we believe that the models provide at least good approximations to possible subject strategies. We also assume in all but the null model that processing of novel information takes longer than processing of familiar information, in that novel information involves temporary overwriting of a subject's previous knowledge. As we will show, this assumption is amply supported by the data in Table 8. We consider each of these models in turn.

<u>Null Model</u>

<u>Model 0</u>. Model 0 would be the most plausible model of information processing if the precueing manipulation failed. According to this model, the precueing manipulation has no effect unless the precue is novel relevant, because only in this condition is the correct answer changed by the information in the precue. Hence, it is possible that the results would show a constant effect of the precue unless there is actually some need of it to compute a new answer. This model predicts that novel relevant (NR) items will be harder than all the others, and that these others--familiar relevant (FR), familiar irrelevant (FI), and novel irrelevant (NI)--will be equal in difficulty:

0. FR = FI = NI < NR

<u>Additive Models</u>

<u>Model 1: Relevant < Irrelevant</u>. In Model 1, there are separate time charges for two elements of information processing: dealing with novelty and dealing with irrelevant information. The motivating notion is that during item solution, subjects are delayed by having to take account of novel information, and also by having to deal with irrelevant information, which is recognized as irrelevant only after a series of successive failed self-terminating relevance tests. Relevance is determined by checking whether the conceptual relation in the precue matches that in the item of the problem. Irrelevance takes longer than relevance because subjects discontinue the relevance tests just as soon as they see that the precue information is relevant to solving the given problem. Furthermore, to the extent that relevant precues act as hints, they may actually facilitate problem solution.

Model 1 can be subdivided into two submodels, 1a and 1b. In 1a, it is presumed that the time charge for novelty is greater than for irrelevance. In 1b, it is presumed that the time charge for irrelevance is greater than for novelty. In sum, the response-time and error-rate predictions of Models 1a and 1b for the four precue conditions--familiar relevant (FR), familiar irrelevant (FI), novel relevant (NR), and novel irrelevant (NI)--are:

1a. FR < FI < NR < NI

1b. FR < NR < FI < NI

<u>Model 2: Irrelevant < Relevant</u>. Model 2 is like Model 1, except that there is an incremental time charge for <u>relevance</u> rather than for irrelevance. On this model, novel information takes longer to process than familiar information, as in Model 1. But relevant information is presumed to take longer to process than irrelevant information, because only relevant information needs to be incorporated into item solution. If a subject determines information is irrelevant, that information can be disregarded. But if the information is relevant, then it is integrated into the knowledge representation used to solve the item, and this integration plus the subsequent use of the extra information adds a time charge to item processing, as well as increasing the probability of an error, due to the addition of an extra step of information processing.

Model 2, like Model 1, can be subdivided into two submodels. In Model 2a, it is presumed that the time charge for novelty is greater than that for relevance, whereas in Model 2b, it is presumed that the time charge for relevance is greater than that for novelty. In sum, the response-time and error-rate predictions are:

2a. FI < FR < NI < NR

2b. FI < NI < FR < NR

Interactive Models

The interactive models differ from the additive models in their assumption that whether relevant or irrelevant information takes longer to process depends upon whether that information is familiar or novel.

Model 3: Relevant < Irrelevant for Familiar Only. In Model 3, as in Models 1 and 2, there is an incremental time charge for novelty. However, it is assumed that for familiar items, successive self-terminating searches result in irrelevant information being slower to incorporate than is relevant information (as in Model 1). Moreover, because familiar relevant information does not change the keyed response on familiar items, such information can immediately be discarded without further processing. On novel relevant items, however, relevant information does change the keyed answer, and hence must be incorporated into solution processing, resulting in an added time charge for relevant information in items with novel precues.

3. FR < FI < NI < NR

Model 4: Relevant < Irrelevant for Novel Only. In Model 4, as in the preceding models, there is an incremental time charge for novelty. However, it is assumed that on items with familiar precues, incorporating precue information into item solution results in these items taking longer to solve than items with familiar irrelevant precues (as in Model 2). On items with novel precues, it is assumed that irrelevant information takes longer to process, because it is more difficult to recognize the precue information as irrelevant. In other words, whereas the relevance of familiar information can be readily assessed, the relevance of information is hard to assess when the information is novel in the first place; hence, the delay for irrelevant information when processed in the context of a novel precue.

4. FI < FR < NR < NI

Results

Reliability of Data

A fundamental, preliminary issue that needs to be addressed before subsequent data analysis is whether the data for the experimental task are reliable. The task is a rather unusual one, and it is possible that, as a result, the data will lack internal consistency, either with respect to items (each item measuring a different ability) or with respect to subjects (each subject employing a different strategy). Item coefficient-alpha reliabilities are equivalent to reliabilities for all possible split halves of items, whereas subject coefficient-alpha reliabilities are equivalent to reliabilities for all possible split halves of subjects. These internal-consistency reliabilities are shown in Table 7. All of the reliabilities are in the high .80s and low .90s, which are highly satisfactory for data-analytic purposes.

Insert Table 7 about here

Correlations between Response Times and Error Rates

A second fundamental and preliminary issue concerns correlations between response times and error rates. This correlation was .62 (p <

.001) across subjects, indicating that there was no speed-accuracy tradeoff, but rather, that subjects who were faster were also more accurate. The correlation was .21 ($p < .001$) across items, indicating that items that took longer to solve were also more susceptible to error. In sum, we need concern ourselves no further either with speed-accuracy tradeoff nor with the possibility that errors tended to be "quick" ones due to subjects giving up without fully attempting item solution.

Basic Statistics

Response times and error rates. Basic statistics for the various experimental conditions are shown in Table 8. Means are shown for both correct response times and error rates as a function of item type (analogies, classifications, series), cueing condition (cued or uncued), and type of precue information (familiar relevant, familiar irrelevant, novel relevant, novel irrelevant). (Patterns of results are essentially identical if all response times, including error ones, are analyzed.)

Insert Table 8 about here

Table 9 shows an analysis of variance upon these means. The results are clearcut: All of the main effects and interactions are statistically significant for the response times, and all of the main effects and all but one of the interactions are statistically significant for error rates. These results indicate that the experimental manipulations affected performance, but more importantly, that these effects were interactive, as would be predicted from either Model 3 or 4. For the main effects, series completions take longer to solve than classifications, which in turn take longer to solve than analogies. Items with novel precues take longer to solve than items with familiar precues, and irrelevant precues take longer than relevant ones. ANOVAs were also calculated for each task, and they were essentially the same as the overall ANOVAs.

Insert Table 9 about here

The various main effects and interactions can be better understood in the context of Table 8, which includes mean response times and error rates on the various item types for each of the four conditions of precueing. Reaction times and error rates show identical patterns: Overall means, as well as the means for the analogy and classifications tasks, clearly support Model 3, the interactive model in which irrelevance adds to response times and error rates for items with familiar precues, but in which relevance adds to response times and error rates for items with novel precues. In the series completions, the response times and error rates for novel relevant items are relatively low, resulting in these means supporting Model 1b rather than Model 3. Thus, the individual task results and the overall means for response times and error rates support Model 3, except for the series completions task, where the results support Model 1b.

Effect of Precues

So far in our models we have only considered simple scores. It is also informative, however, to look at the pattern of differences between cued and uncued scores in order to examine the effect of cues on problem solution (Table 10). The expected, obvious patterns occur clearly: that the difference scores for items with irrelevant precues and/or novel precues are all positive; that is, both novel and irrelevant precued items have a higher response time and error rate than the same items uncued. It is most interesting, however, to look at the scores for items having familiar relevant precues, because solution of these items is sometimes hindered and sometimes facilitated by the presence of precues.

Insert Table 10 about here

The familiar relevant precues make analogies more difficult while facilitating solution of classifications and series completions. There are two possible reasons for this effect. The first reason relates to the difficulty of the type of item. As can readily be seen from Table 8, analogies are clearly the easiest uncued items. This is probably because solving analogies is a well-practiced skill in the population sampled (Yale undergraduates). The analogies in this experiment, which employ only common words and concepts, can be solved almost automatically by these subjects, whereas the classifications and series completions are less familiar and hence more difficult. The effect of a precue--even a familiar relevant one--on analogy solving is to cause an interruption of processes that would otherwise proceed automatically. Even if the precue gives helpful information, incorporating it expends more time than it saves. On the other hand, the precues for classifications and series completions are potentially more helpful because the subjects need more help on the more difficult items.

The second possible reason for the differential helpfulness of familiar relevant cues is that the more flexible the structure of an item, the more difficult it is to determine the relevance of a precue. Analogies are highly flexible in that there can be a wide variety of relations between terms. Consider, for example, the relationships between BAGEL and DOUGHNUT, between HERO and ADMIRATION, and between EYE and BLINDING (all from test items). In classifications, the relations between terms are more constrained; in an item, all terms are the same part of speech, and they are all related by group membership. In series problems, the relationships between terms are even further restricted; in addition to satisfying the constraints mentioned above for classifications, the terms must also be arranged in a progression. Thus, in analogies it should be fairly difficult to determine whether or not a precue contains relevant information. In other words, nearly any precue which mentions any aspect of any of the analogy terms could conceivably be relevant to problem solution.

In classifications, because the terms must be related by group membership, it should be fairly easy to spot a relevant precue, and a relevant precue should be fairly helpful, in that it gives away the rule of to what group the terms belong. The precue in series problems is even easier to recognize as relevant and is more helpful, in that if it is relevant, it really "gives away" the answer, and if not, it is readily perceived as unhelpful, and can be quickly discarded.

According to the above rationale, the following would be expected: For analogies, both relevant and irrelevant cues are unhelpful and time-consuming, in that they mainly distract, whereas for classifications and series completions, relevant precues are helpful, and it is less difficult to recognize a precue as relevant. These expectations do describe the pattern of results obtained for items with familiar precues. As shown by both response times and error rates, solution of analogies is impeded by precues, whereas solution of classifications and series completions is facilitated by precues.

Similar patterns hold for the novel precues for analogies and series completions; however, the results for novel classifications are anomalous. The reason seems to relate to the actual novel relevant cues used in the classifications. The most obvious type of novel relevant precue which could have been constructed for classification problems turned out to be trivially easy, such as:

Celery is a fruit.
apple, pear, peach, grapes,
1. orchard
2. produce
3. celery
4. jam

In order to avoid such items, which merely entail substitution rather than reasoning, the items were constructed as follows:

Fruits have long green stalks.
apple, pear, peach, grapes,
1. orchard
2. produce
3. celery
4. jam

Thus, in the novel precues for classifications (unlike those for analogies and series completions), the change induced by the precue is a change in the properties of the words in the stem, not in the word which completes the item. The effect of this extra inductive step was to make the novel relevant precues more difficult to reason with for classifications than for either analogies or series completions. Consequently, the novel relevant classifications have comparatively high reaction times and error rates.

In summary, the effect of precues seems to be as follows: Precues hinder item solution if the precue is novel and/or irrelevant, if the item is easy and automatized in the uncued state, if the structure of the item is flexible enough that it is difficult to determine whether or not the precue is relevant, or if the precue is nonobvious by being several inductive steps away from the problem stem. Conversely, a precue can facilitate item solution if the precue is familiar relevant, the item is fairly difficult to start with, the item is of a rigidly structured type where the precue can be readily recognized as helpful, and the precue is inductively close to the item stem.

Correlations of Experimental Task with Psychometric Test Scores

Table 11 shows correlations between experimental task and psychometric test scores. Correlations are shown both for the individual tests, and for approximation factor scores based on a varimax-rotated principal-axis factor analysis of the data. The factor analysis yielded two interpretable factors with eigenvalues greater than 1, namely, a reasoning factor (DAT Verbal Reasoning, Cattell Culture-Fair Test of g, Insight Test) and a verbal/perceptual factor (Crossing Out A's, Vocabulary).

Insert Table 11 about here

These correlations address three basic questions. The first question concerns convergent validation, whereas the second and third questions concern discriminant validation. First, are scores on the experimental tasks related to scores on the psychometric tests? One would expect most of the correlations to be statistically significant, but moderate, as the abilities tapped by the experimental tasks should be related but nonidentical to those tapped by the psychometric tests (see Hunt, Frost, &

Lunneborg, 1973, for similar logic). Second, do the correlations with the reasoning tests (and factor) differ from those with the verbal/perceptual factor? In particular, the reasoning tests were chosen to measure fluid abilities similar to those tapped by the experimental tasks, whereas the verbal and perceptual tasks were chosen to measure crystallized abilities different from those tapped by the experimental tasks (see Hunt, Lunneborg, & Lewis, 1975, for an information-processing analysis of verbal, or crystallized abilities). Third, are the correlations for the uncued and cued conditions with the psychometric tests different? The triarchic theory predicts that the more nonentrenched (precued) items should be the better measures of intelligence, and the psychometric tests of fluid ability were chosen for being among the best measures of fluid reasoning ability. Hence, the correlations for the precued items might be expected to be higher than those for the uncued items.

The answers to these questions are fairly straightforward. First, most of the correlations in the table are statistically significant, and generally low to moderate in magnitude. Second, correlations of the experimental tasks with the reasoning tasks are clearly higher than those with the verbal/perceptual tasks. For the factor scores, for example, all of the correlations of reaction times and of error rates with the reasoning factor (ranging from -.31 to -.46) are statistically significant, whereas none of the correlations with the verbal/perceptual factor (ranging from -.05 to -.17) are significant. Thus, the experimental tasks do appear to tap abilities related to those tapped by the psychometric tests. Third, six of six comparisons between correlations of uncued versus precued response times and error rates with the psychometric reasoning tests come out higher for the precued than for the uncued items. In sum, the experimental tasks showed the predicted patterns of convergent-discriminant validity with respect to the psychometric tests.

Discussion

This experiment investigated information processing during the solution of analogies, classifications, and series completions either taking the standard form or else preceded by precueing information that could be novel or familiar, and relevant or irrelevant. The data analyses investigated both the internal and external validities of the tasks.

With respect to internal validation, it was found that for analogies and classifications, subjects take longer to process irrelevant than relevant information if the precue is familiar, but they take longer to process relevant than irrelevant information if the precue is novel. For series completions, however, both novelty and irrelevance add time to the processing of information, with the time for irrelevance greater than that for novelty. Figure 1 shows a general information-processing model—formalized as a flow chart—that provides a plausible account of information processing in all three tasks. The tasks differ only in the proportion of time the various steps take, which is a factor of item type, cue difficulty, and so on. These variables affect what specific information-processing model applies to a given task. For instance, as previously discussed, applying novel relevant precues is easy for series completions relative to the other two tasks, so the resulting model is additive rather than interactive.

Insert Figure 1 about here

In the general model shown in the flow chart (which applies to precued items only—for standard models, see Sternberg & Gardner, 1983), subjects first read the precue. If the precue is familiar, then the subjects access category information needed for problem solution. Subjects then self-terminating as soon as they find how the information is relevant to item solution. At this point, they incorporate the precue if it is relevant, solve the problem, and then respond (see Sternberg & Gardner, 1983, for how similar uncued problems are solved). If successive tests fail to show the relevance of the precue, subjects discontinue relevance checking, reject the precue as irrelevant, solve the problem, and respond.

If the precue is novel rather than familiar, then subjects have to create new category information for the counterfactual. After requesting the item stem, they then do a series of relevance checks for the novel precue. If, after a series of checks, they are unable to see any relevance for the precue, they reject the cue, solve the problem, and respond. If, however, the precue is relevant, then they combine the precue with the stem, a nontrivial process since the precue not only overrides their previous knowledge but also changes the answer to the problem. After combining the cue with the stem, the subjects solve the problem and respond.

With respect to external validation, it was found that the nonentrenched induction tasks overlapped with psychometric tests in terms of abilities measured, that the abilities measured were fluid rather than crystallized, and that the precued (more nonentrenched) items were better measures of fluid abilities than were the uncued ones. The nonentrenched induction tasks thus fulfill their original goal of being essentially IQ-test-like in their surface structure, in that they can be answered rather quickly with an unambiguous keyed answer. But they are unlike standard items on intelligence tests in being more nonentrenched--or unusual--in their information-processing requirements.

The results of this experiment are consistent with the notions of Raaheim (1974), Snow (1980), and Sternberg (1982, 1985) that relatively nonentrenched tasks provide particularly apt measures of intelligence, and particularly, of fluid intelligence. The ability to cope with relative novelty is an important aspect of intelligence, and it can be measured efficiently and in a theoretically-based way using the nonentrenched analogies, classifications, and series completions used in the present research.

Expertise and Flexibility: The Costs of Expertise[3]

During the past three decades, research efforts directed toward understanding the nature of expertise have increased tremendously. Differences between experts and nonexperts in complex problem-solving domains have been shown to be both quantitative and qualitative. Not only do experts perform better than do nonexperts on quantitative measures, but they also seem to structure problem representations differently and appear to apply different strategies. Across a variety of domains, it is found that experts tend to conceptualize domain-related problems in abstract terms, whereas nonexperts apparently rely more on surface-level features.

For example, in their classic study on expert chess players, Chase and Simon (1973), replicating deGroot's (1965) earlier findings, reported that more experienced players spent less time, made fewer errors, needed fewer glances, and took in more information per glance than did less experienced players when their task consisted of memorizing or reproducing briefly presented, meaningful chess patterns. However, experts' encoding ability dropped to the level of a novice when the experts were forced to deal with meaningless chess patterns. Examining the recall clusters of their players, Chase and Simon found that the clusters of experienced players frequently were based upon attack or defense configurations, implying an abstract knowledge representation.

Chi, Glaser, and Rees (1981) extended this result to physicists on the basis of conceptual sorting tasks and verbal protocols. They suggested that experts' problem representations tend to focus on abstract physical principles, whereas novices represent problems in terms of surface features. Researchers investigating different skill domains have arrived consistently at the same conclusion (e.g., Adelson, 1981; 1984; Charness, 1979; Kay & Black, 1985; Reitman, 1976; Schoenfeld & Herrmann, 1982).

Studies focusing on the structure of knowledge representations generally attempt to understand qualitative changes that arise with growing expertise in a field. These studies, however, do not necessarily address the issue of quantitative changes in the cognitive system that also go along with increasing expertise and that lead to remarkable decreases in speed of performance. For example, in a study of expert and nonexpert bridge players, Charness (1979) showed that experts needed less time to perform qualitatively better than did novices in four different, bridge-related tasks: planning the play of a contract, rapid bidding, incidental learning, and recall of briefly exposed, meaningful bridge hands. The only task in which experts did not perform faster and better than did nonexperts was the recall of briefly presented, unstructured hands.

Several researchers have proposed theories that address why experts can solve problems more quickly than nonexperts can (e.g., Cheng, 1985; Kahneman, 1973). One of the most elaborated among those theories was introduced by Anderson (1982). Based upon the mechanisms of his ACT* model of cognitive architecture, Anderson proposed that expertise develops from an earlier declarative stage, in which general interpretative procedures are applied to facts about the skill domain, to a procedural stage in which the domain knowledge is compiled into procedures. According to his view, the mastery of a cognitive skill is never completely finished. Even compiled procedures continue to be fine-tuned by means of generalization, discrimination, and strengthening.

Taken together, the above findings suggest that the degree of proceduralization and the abstract nature of the knowledge organization are among the most important ingredients of skilled problem-solving. Proceduralized domain-specific problem-solving strategies operating on problem-suited knowledge representations give experts both a speed and a quality advantage over nonexperts.

Generally, studies on expertise have focused on understanding why experts do better than nonexperts. Theories of expertise, such as the ones described above, have been constructed to account for experts' superior problem-solving ability, thereby emphasizing the benefits of expertise. Such theories, however, should not be restricted to situations in which experts outperform nonexperts. Complete theories of expertise should be able to predict under which circumstances experts will not perform better than nonexperts or will even be outperformed by nonexperts. So far, however, theories have only occasionally been linked to the costs of expertise (but see Adelson, 1984).

One reason why experts sometimes might be outperformed by nonexperts could derive from the inflexibility of their information-processing system. We hypothesize that a cognitive system that has been developed specifically to perform within the boundaries of a well-specified problem area need not necessarily be very flexible. This inflexibility might reveal itself in one or both of two different ways. First, the system could produce negative transfer as a result of being unable to inhibit old solution strategies when response requirements change. Thus, an expert physicist, for example, might find it very difficult to inhibit older, proceduralized explanation patterns to account for familiar problems even though she is aware of new findings that indicate the inadequacy of such solutions. Second, the cognitive system's knowledge base might be hard to modify. To stay with our example: Even if our physicist had no problems inhibiting old solutions, she might still be hard pressed to incorporate the new findings into her existing knowledge base.

Inhibition and adaptation, the two components of flexibility, have both been used as explanatory mechanisms in the long controversy surrounding theories of associative learning. Different frameworks explaining the negative interference effects obtained in the basic A-B, A-C paradigm have tended to emphasize one or the other of the two different aspects of flexibility. Whereas proactive interference (e.g.,Underwood, 1957) and list discrimination accounts (e.g., Sternberg & Bower, 1974) described interference effects in terms of subjects' inability to inhibit older, previously established associations, organizational theory (e.g., Tulving, 1966) tried to link them to adaptation difficulties.

In an attempt to understand why people find it difficult to reason with new concepts, Tetewsky and Sternberg (1986) showed that subjects experience more difficulties in novel situations when concepts are partly unfamiliar than when they are completely unfamiliar. Tetewsky and Sternberg refer to adaptation difficulty and inhibition difficulty as two possible explanations for the result they obtained.

We claim that inflexibility and expertise are inextricably linked. A problem solver's inability to inhibit existing response patterns and to modify old knowledge when response requirements change (i.e., inflexibility) are affected by his knowledge organization and by its degree of proceduralization (i.e., expertise). A highly structured system that has proceduralized its processing of knowledge will be relatively inflexible.

In Experiment 1 we test the hypotheses that (a) the structure of a knowledge base and (b) its degree of proceduralization affect the flexibility with which a problem solver can respond to novel task demands. We expect that problem solvers who have proceduralized their solution strategies to a great extent will be more affected by changes in task demands than will problem solvers who respond to problems in a less proceduralized manner. Thus, experts in any given domain will generally be more vulnerable to task demand modifications than nonexperts will be. Furthermore, we assume that the effects of changes in task demands are larger for those demands that are incompatible with the structure of the knowledge base than for demands that are compatible. Therefore, experts will be more affected when new task demands call for deep, abstract principles to be changed than when surface features are to be changed. For nonexperts, the reverse is expected. That is, they will be more affected by surface changes than by deep, abstract changes.

The specific domain of knowledge chosen to test these ideas is the game of bridge. (Appendix A describes the basic bridge rules for readers unfamiliar with the game.) There have been only few studies on skill differences in bridge (e.g., Charness, 1979; Engle & Bukstel, 1978). The few studies conducted generally obtained results that are compatible with the findings reported in other domains. It appears that master bridge players encode abstract distributional features of bridge hands that automatically evoke strategies containing long lines of play, like "play an end game." Nonexperts, on the other hand, seem to encode hands primarily in terms of surface features (honor-cards, for instance), that are associated with smaller lines of play, like "take a trick."

In the first experiment, bridge players of varying levels of skill played 12 simulated bridge games on a computer. Half of the games were played under normal conditions. In the other half, players were instructed to play slightly different versions of bridge. Version 1 introduced new nonsense names for honor-cards and suits; version 2 rearranged the order of honor-cards and suits; and version 3, the lead-rule change, modified the rule determining who began each play. Instead of the player who won the last trick, which is the common rule in bridge, the player with the lowest card in the last trick led into the next trick. The different versions were intended to tap different levels of subjects' information processing. Versions 1 and 2 were considered surface modifications; version 3 was expected to exert its effect on a deeper, more abstract and strategic level.

Experiment 1

Experiment 1 tested three different hypotheses. First, experts will play bridge more quickly and qualitatively better than nonexperts will when the original bridge rules are in effect. Second, experts will be more affected by the three rule modifications than nonexperts will be. Finally, experts will be more affected by the deep-structural change than by the surface-structural changes, whereas for nonexperts, the reverse will be true.

Method

Subjects

Thirty-four bridge players (16 males and 18 females) from the New Haven area were recruited via advertisements in three local duplicate bridge clubs. The president of each club assigned each player to one of 10 categories according to level of skill. The category system that was used is shown in Appendix B. Expertise scores for each subject were obtained by averaging across the three raters' assignments. The players ranged in age from 21 to 70. They were run in individual sessions and were paid $5 per hour for their participation.

Materials

Experimental task. The bridge program used in this study was based on Borland International's Turbo Pascal version of Turbo Bridge and was modified by Peter Frensch to suit the purposes of the study. It was run on an IBM PC microcomputer.

Hand distributions of all games were randomly chosen by Peter Frensch prior to data collection (out of a pool of computer-generated random games) and were displayed on the screen using the same abstract representation as is commonly found in bridge books and newspaper columns. The screen display was identical for the bidding and playing phases of a game. Figure 2 shows a hypothetical screen configuration as it might have appeared in the first trick of a game.

Insert Figure 2 about here

As in an actual bridge game, players saw only their own hands during the bidding. When the contract was determined and the opening lead played, dummy's hand was made visible and remained on the screen for the rest of the game. Bids made and cards played in a trick were displayed within the rectangular "bridge table" and, additionally, in the order they were played, in the lower left corner of the screen. The number of tricks won was updated after each trick and separately displayed for teams West/East and North/South. Subjects either played hand South or, when they won the contract or were dummy, hands South and North with the remaining hands being played by the computer. Information about the program's activities was continuously displayed (e.g., "North bidding," "West playing"). Subjects were prompted to make a bid or to select a card (e.g.,"South: ?"). Card and suit ranks (bid levels and bid suits during the bidding phase) were displayed together with the prompt. Special keys on the computer keyboard were assigned to cards, bid levels, and suits.

Three different modifications of the regular bridge rules were used: a name change, a rank-order change, and a lead change. In the name-change condition, the familiar card and suit names (ace, king, queen, jack, and spade, heart, diamond, club) were replaced by nonfamiliar, nonsense names (rutz, lork, dill, beib for cards, and pular, biref, ramog, kamer for suits, respectively). The rank-order change modified the familiar card and suit rank orders (ace, king, queen, jack to king, jack, ace, queen, respectively, for cards, and spade, heart, diamond, club to diamond, heart, club, spade, respectively, for suits). The lead change modified who began each play. Instead of the player who won the last trick, the player with the lowest card in the last trick led into the next trick.

Questionnaire. In addition to participating in the experimental task, subjects were given a questionnaire asking their age, number of years they have played bridge, number of master points accumulated in total and during the past two years, frequency of duplicate and rubber bridge play per month, an estimate of their level of bridge skill on a 1-9 scale (from poor to excellent), and their experience with other card games. (Master points in bridge are awarded by the American Contract Bridge League to those scoring in the top positions at duplicate bridge tournaments. The number of points varies with the number of participants and type of event.)

Design

The dependent variables were play time, number of tricks won per game, and number of games won. Independent between-subjects variables were level of expertise (nonexpert, expert) and type of rule modification (name change, rank-order change, lead change). Game (1 to 12) was a within-subjects variable. In each expertise group, subjects were randomly assigned to one of the three types of game modifications.

Procedure

Approximately half of the subjects in each of the two expertise groups answered the questionnaire and took part in two different bridge studies (Experiments 2 and 3 of this report) at the first session and completed the experimental task in a second session; for the other half, the order was reversed. The two sessions took place on two different days.

Instructions for the questionnaire were administered orally. Presentation of the questions and recording of the answers were performed by the computer.

Instructions for the experimental task were administered via computer. Subjects were told that they would play 12 bridge games on the computer. They were shown how to enter bids and plays and were familiarized with the bidding system, which was simplified Goren. Subjects were also informed that there was no carry-over in terms of vulnerability from one game to the next. They were told that they would always play hand South and that, when South was declarer or dummy, the computer would expect them also to play hand North.

Each game began with the bidding. When declarer and dummy were determined, the player to the left (counterclockwise) of the dummy played the opening lead. Then the dummy cards were revealed. When each trick was completed, the cards played remained on the "table" for further analysis until the subject hit a key. The player who won the previous trick led into the next one (except in the lead-change condition). When the subject had finished playing all 13 tricks, the next game was started.

Each subject received the same 12 games in different randomized orders. Games were blocked with regard to type of game modification. Games 1 through 3 and 10 through 12 were regular, unchanged games; games 4 through 9 were played with a modification of the familiar bridge rules (name change, rank-order change, or lead change). Each player was exposed to only one form of rule modification. To familiarize themselves with the system, subjects played two practice games prior to the experimental task.

When they had completed three experimental games, players were presented with the new rule embedded in a cover story. The cover story was identical for all new rules, telling subjects to imagine themselves traveling in a foreign country where, one evening, they were invited to a card game. Predictably enough, these foreigners played a game that was very similar to bridge with only slight modifications. Appendix C gives the complete cover story for the modification of card and suit names. In the name-change and rank-order-change conditions, the experimenter changed the key labels on the computer keyboard to reflect the new names or rank orders.

Subjects were allowed to take as much time as they needed to select a bid or a play. Plays, play times, the number of tricks won, and whether the player had won or lost the game were recorded by the computer.

Results

Group Assignment

Two expertise groups (experts and nonexperts) were formed on the basis of subjects' assignments to the expertise category system. Subjects were rated by the three presidents of the clubs we had contacted. Scores for each subject were obtained by averaging across raters' assignments. Interrater reliabilities were .82, .85, and .89. As can be seen in Table 12, expertise (exprts) was positively correlated with self-reported level of skill (Estlvl), total number of master points accumulated (MPtot), number of master points accumulated during the past two years (MPlast), and number of years played (Yrsplyd), but not with age. Subjects with average expertise scores of 1-4 were assigned to the nonexpert group; those with scores 5 and higher formed the expert group. The particular cutoff point for the two expertise categories was not chosen on the basis of any conceptual meaning but, instead, was dictated solely by the available data pool (median cut).

Dichotomization was a necessary precondition to analyze data by use of analysis of variance. (Simple regression methods would not have required such a categorization but would have been much more inconvenient to use.) Table 13 provides a snapshot of expert and nonexpert characteristics. As can be seen, groups differed on expertise scores, on self-estimated level of skill, on the number of master points they had collected in total and during the past 2 years, and on the number of years they had played bridge prior to testing. The two groups did not differ in age.

Insert Tables 12 and 13 about here

Reliabilities of Dependent Variables

An important issue that we wanted to address prior to subsequent data analysis is the reliability of the data. Because the experimental task is a rather unusual one, the obtained data might have lacked internal consistency. Because the specific game situations to which subjects responded were not constant across subjects, but were dependent on earlier plays, split-half reliabilities, corrected by the Spearman-Brown formula, were computed for all dependent variables. As Table 14 shows, the reliabilities for play time were in the .80s and low .90s, which is generally considered to be highly satisfactory for data analysis. Data for the number of tricks won per game and for the number of games won generally had lower reliabilities, particularly when unchanged and modified games were considered separately. This is to be expected, given the categorical nature of the two variables and the differences in strategy that subjects of different skill level employ to deal with the rule modifications.

Insert Table 14 about here

Correlations among Dependent Variables

A second, preliminary issue concerns the correlations among dependent variables. Of particular interest is the correlation between speed and quality of subjects' responses. Because of the nature of this study, error rates could not be obtained. Instead, number of tricks won per game and number of games won were used as measures of subjects' qualitative task performance. As can be seen in Table 15, play time was only moderately correlated with the two qualitative measures. This was true for both unchanged and changed games. It should be noted that the two overall correlations included all observations and, thus, did not suffer from the low reliabilities reported for the qualitative measures, especially in the modified games. We may conclude, therefore, that subjects did not sacrifice quality for speed of response.

The overall correlation between number of tricks won per game and number of games won was .36, $p < .001$, indicating that subjects who won more tricks also were more likely to win more games than subjects who won fewer tricks. Somewhat surprisingly, however, this correlation is far from perfect. Winning many tricks, it appears, does not necessarily make for a good player, that is, a player who wins many games.

Insert Table 15 about here

Basic Statistics

Prior to inferential data analysis, the within-subjects factor of game (1 to 12) was broken down into four blocks of three games each. Thus, blocks 1 and 4 contained only games that were played under the normal bridge rules; blocks 2 and 3 were games played under modified rules. Tables 16 and 17 give the basic statistics for all dependent variables in the various conditions.

A first, informal inspection of the data revealed that subjects' play times in all experimental conditions were faster in block 4 than in any other block. Similarly, subjects won fewer tricks and fewer games in block 4 than in any other block. Block 4 was originally designed to test whether subjects would display any interference effects when they had to return to the original bridge rules. Because we did not use control groups that would have allowed us to assess and partial out practice effects, any effect of retroactive interference could have only been tested if block 4 performance was slower than block 1 performance. This was not the case, however. One possible reason for this finding might have been subjects' decreasing attention span at the end of the task due to the length of the study. (Average total times spent on the experimental task were 2 hours, 59 minutes for nonexperts and 2 hours, 55 minutes for experts.) Because the attention-deficit hypothesis could not be ruled out, data analysis was limited to subjects' performances in blocks 1 to 3.

Insert Tables 16 and 17 about here

Base-line Performance (Regular rules)

The first hypothesis predicted that experts would play faster and qualitatively better than nonexperts when regular bridge rules apply. Basic statistics for experts' and nonexperts' performances in the unchanged games (block 1) are shown in Table 18. Because the dependent variables were correlated, a multivariate analysis of variance was performed. For the main effect of expertise, Wilks $L=0.648$, $F(1,28)=4.71$, $p<.01$. The effect accounted for approximately 35.3% of the total variance. Experts played faster and won more games, but won fewer tricks per game than did nonexperts. To decide which of the dependent variables contributed most toward the discrimination of the expertise groups, a follow-up discriminant analysis was performed. The structure coefficients of the discriminant function were 0.74, 0.41, and -0.51 for the dependent variables of play time, number of tricks won, and number of games won, respectively, indicating that the two expertise groups differed most with regard to the time they took to select a play.

Insert Table 18 about here

Effects of Rule Modifications

The second and third hypotheses predicted that experts would generally be more affected by the rule modifications than nonexperts would be, and that they would be more affected by the deep-structural rule change than by the two surface-structural changes. We expected the latter pattern to be exactly reversed for nonexperts. Thus, nonexperts should be more affected by the surface-structural changes than by the deep-structural change.

To assess changes in the dependent variables that were due to the rule modifications, percentage scores were computed by dividing each subject's performance in the changed games (averaged across blocks 2 and 3) by his or her performance in the regular games (block 1). We computed three percentage scores for each subject, one for each dependent variable. Table 19 gives the mean percentage increase or decrease for the dependent variables in all experimental conditions. A multivariate analysis of variance on the percentage scores revealed a significant main effect of expertise, Wilks $L=0.669$, $F(1,28)=4.30$, $p<.05$, accounting for approximately 25.3% of the total variance. A contrast, testing the interaction hypothesis, was also significant, Wilks $L=0.699$, $F(1,28)=3.74$, $p<.05$. The contrast accounted for approximately 45.1% of the total variance. The main effect of rule modification and the (overall) interaction of expertise with rule modification were both nonsignificant, $p>.10$.

Looking in more detail at the differential effects of the rule modifications on the two expertise groups, we found that experts' play times increased by more than did nonexperts' as the new rules were introduced. Also, the quality of experts' play, as described by the number of games won, was more negatively affected than was nonexperts'. In fact, the nonexpert group won more games, on average, when playing under the new rules than when playing under the original rules. However, the nonexpert group won fewer tricks when the rule changes were introduced, whereas the experts managed to increase the number of tricks they won despite the changes. Increasing familiarity with the experimental setting and with the way the computer played might have been responsible for subjects' gains in the qualitative measures that were observed, despite the introduction of new task demands.

With regard to the significant interaction contrast, we found that the hypothesized pattern of results was obtained for each of the three dependent variables individually. As Table 19 shows, experts were more affected by the deep-structural change than by the surface-structural changes in all measures, whereas nonexperts were more affected by the surface-structural changes than by the deep-structural change, a result that, again, showed up in all dependent measures.

To assess which of the dependent variables were primarily responsible for the significant results, we conducted two follow-up discriminant analyses. The structure coefficients for the canonical function that separated the two expertise groups were 0.75, 0.55, and -0.25 for the dependent variables of play time, number of tricks won per game, and number of games won, respectively, indicating that the differential effect on the two expertise groups showed up most clearly in the time needed to select a play.

The second discriminant analysis (for the interaction contrast) was done separately for experts and nonexperts. For experts, the structure coefficients obtained were 0.97, -0.23, and -0.16 for play time, number of tricks won, and number of games won, respectively. It appears that the difference between surface changes and deep-structural change is primarily a matter of time for experts. In other words, experts need more time to deal with the deep-structural change than to deal with the surface changes. The quality of their play, however, does not seem to be differentially affected by the two types of changes.

For nonexperts, we found structure coefficients of -0.44, 0.21, and 0.66 for the dependent measures of play time, number of tricks won, and number of games won, respectively. For less experienced bridge players, therefore, the difference between the types of rule changes is associated with a difference in quality of play. Nonexperts do worse on the surface changes primarily with regard to the quality of their performance.

Insert Table 19 about here

<u>Adaptation to Rule Modifications</u>

In addition to assessing the overall impact of the different rule modifications on the two expertise groups, we wanted to examine how well the subjects performed the new tasks after they had time to adapt to the changes. One way of addressing this question is to compare subjects' performances on the experimental blocks 1 and 3 only. We therefore computed new percentage scores for each subject by taking the ratios of block 3 performance and block 1 performance on all dependent measures. A multivariate analysis of variance on these scores revealed no significant effects, indicating that both expertise groups had adapted quite well to all three rule changes. In fact, experts appeared to have won back their original speed and quality advantage over nonexperts in the two surface-change conditions. Excluding the abstract-change condition, we obtained a marginal main effect of expertise, Wilks $L=0.791$, $F(1,20)=2.29$, $p<.10$, for the two surface-change conditions. Block 3 means for these conditions are shown in Table 20, separately for the two expertise groups. When compared with the original performances in the unchanged games (compare Table 18), it appears that both expertise groups, after only a few practice games under the two surface rules, have "recovered" already in terms of play time and number of tricks won. The only variable on which both groups did not reach their original level of performance was the number of games won.

Insert Table 20 about here

Discussion

Experiment 1 was conducted to clarify the relation between expertise and flexibility. We had hypothesized that two variables, namely, the organization of the knowledge base and the degree of proceduralization, would affect the flexibility of a problem solver. The main results of this study can be summarized as follows. First, changes in task demands generally affect subjects who have highly proceduralized their problem-solving strategies more than those who rely on less proceduralized strategies. Experts are generally more affected by the rule modifications than nonexperts are. Second, changes in task demands that are incompatible with the structure of the knowledge representation have more pronounced effects on subjects' problem-solving processes than have changes that are compatible. Experts were more affected by the deep-structural change than by the two surface changes, whereas nonexperts were most affected by the surface changes. And finally, both expertise groups seem to be able to adapt their existing knowledge bases quite rapidly. In all conditions, with the exception of the deep-structural-rule-change condition for experts, subjects reached approximately their original performance levels after only a few games.

It is interesting to note that the difference between surface and abstract changes for expert bridge players appears to be most pronounced in the play-time variable, whereas for nonexpert players, the difference between these two types of rule modifications seems to manifest itself most clearly in a quality measure, namely, the number of games won. Apparently, experts use the additonal time they need when dealing with the abstract change to employ new and effective game strategies. Thus, their inflexibility might reflect a difficulty in adapting their existing knowledge base, rather than an inability to inhibit older solution patterns.

Nonexperts, on the other hand, seem to perform particularly poorly in the rank-order-change condition, indicating that their inflexibility might be primarily due to their inability to suppress older responses.

Generally, Experiment 1 provides moderate support for the hypotheses that the degree of proceduralization and the structure of the knowledge base affect the inflexibility of a cognitive system. One might object that it is not sufficient to demonstrate that subject-groups that organize their knowledge bases differently and rely on different degrees of proceduralization also differ in how inflexible they are, rather that, in order to claim strong support for our hypotheses, we would have to show that subject-groups that do not differ on the relevant variables are also equally flexible (or inflexible). Study 2 was designed to test the latter prediction.

Experiment 2

In the second experiment, 34 bridge players of different levels of skill were instructed to generate opening bids to given bridge hands as quickly as possible. As Charness (1979) pointed out, even novice bridge players are able to generate a reasonable opening bid. Furthermore, the choice of opening bids is based primarily on the distribution of honor-cards and on the number of total cards per suit in a given hand. Although there are a large number of different bidding systems to determine a bid, these systems are based upon the same properties of hands and, generally, do not arrive at different opening bids. The selection of an opening bid is a rather complex computational process that, for most players, does not involve any strategic considerations. With regard to opening bids, more and less experienced bridge players are not expected to differ in their classification of hands into bid categories. Thus, expert and nonexpert bridge players do not use differently structured knowledge bases. In addition, because of the absence of strategic considerations, the computational process of selecting an opening bid can be expected to be proceduralized rather quickly.

Experiment 2 tested three hypotheses. First, experts and nonexperts will select identical opening bids in the same amount of time when the original bridge rules are in effect. Second, experts will generally not be more affected by the rule changes than will nonexperts. And third, experts and nonexperts will not be differentially affected by the rule modifications. In other words, there will be no significant interaction between expertise and type of rule change. Because the selection of opening bids is based primarily upon surface features of hands, we expect both expertise groups to be more affected by the surface changes than by the deep-structural change.

Method

Subjects

Subjects were the same as in Experiment 1.

Materials and Procedure

Experimental task. For the bidding task, subjects were seated in front of a computer screen and keyboard. They were shown slides with "real" bridge hands on the wall just above the computer screen and were instructed to select an opening bid as quickly as possible by pressing one of the predesigned keys on the computer keyboard. Bridge hands were displayed in the same way players usually hold their cards—with the cards fanned out. Subjects were told that they could use any bidding system they wanted, but that they should use it consistently during the whole session. Slide onset and offset were controlled by the computer. As soon as subjects had responded to a slide, the next slide was automatically displayed and stayed visible until the opening bid was selected.

Subjects responded to 40 slides (bridge hands) in total. Half of these slides were bid under normal unchanged bridge rules; the other 20 slides were bid under slightly modified bridge rules.

Types of rule modifications. Types of rule modifications were the same as those used in Experiment 1. They were: a name change, a rank-order change, and a lead change. In the name-change condition, the familiar card and suit names (ace, king, queen, jack, and spade, heart, diamond, club) were replaced by nonfamiliar, nonsense names (rutz, lork, dill, beib for cards, and pular, biref, ramog, kamer for suits, respectively). The rank-order change modified the familiar card and suit rank orders (ace, king, queen, jack to king, jack, ace, queen, respectively, for cards, and spade, heart, diamond, club to diamond, heart, club, spade, respectively, for suits.) The lead change modified who began each play. Instead of the player who won the last trick, the player with the lowest card in the last trick led into the next trick.

Slides were blocked with regard to the rule modifications. The first 10 and the last 10 slides were bid under normal conditions, the rest under one of the modified conditions. Each subject received only one modification in the task. Bids and reaction times were recorded by the computer.

Design

Independent between-subjects variables in this task were level of expertise (nonexpert, expert) and type of rule modification (name change, rank-order change, lead change). Slide (1 to 40) was a within-subjects variable. In each expertise group, subjects were randomly assigned to one of the three types of game modifications, with the restriction that no subject received the same type of change she had already experienced in Experiment 1.

Results

Reliabilities of Dependent Variable

As Table 21 shows, all of the item and subject reliabilities (Cronbach's alpha) for the dependent variable bid time were in the high .80's and .90's, which is considered very satisfactory for data-analytic purposes.

Insert Table 21 about here

Base-line Performance (Block 1)

Basic statistics for bid times in the various experimental conditions are shown in Table 22. Means are listed as a function of expertise, type of rule modification, and block. As in the previous experiment, the within-subjects factor of slide (1 to 40) was broken down into four blocks of 10 slides each. Blocks 1 and 4 contained only slides that were bid under the normal bridge rules; blocks 2 and 3 contained slides bid under modified rules. Hypothesis 1 predicted that all subjects would choose identical opening bids in the same amount of time. A comparison of experts' and nonexperts' bid times prior to any rule modification (block 1) revealed what we had expected. Although experts bid faster than nonexperts (7.79 sec. for experts, 9.51 sec. for nonexperts, respectively), this difference was not statistically reliable, $F(1,28)=2.73$, $p>.10$. Apparently, all subjects had proceduralized the strategy to select an opening bid to roughly the same extent. Inter-subject agreement with regard to which bid to choose was very high, 90.10% and 79.22% for the nonexpert and expert groups, respectively. In addition, the opening bid most favored by the nonexperts was identical to the one preferred by the experts for all twenty bridge hands that were bid under the original rules. Thus, experts and nonexperts tended to select identical opening bids in approximately the same amount of time.

Insert Table 22 about here

Effects of Rule Modifications

Hypotheses 2 and 3 predicted that experts and nonexperts would generally be equally affected by the rule modifications and that both groups would be more affected by the surface modifications than by the abstract change. As in Experiment 1, where we assessed changes in the dependent variables that were due to the rule modifications, we computed percentage scores by dividing each subject's bid time in the changed games (averaged across blocks 2 and 3) by her bid time in the regular games (averaged across blocks 1 and 4). An analysis of variance on these scores revealed only a significant effect of type of rule modification, $F(2,28)=3.59$, $p<.05$, accounting for 19.64% of the total variance. Figures 3 and 4 show that all subjects were more affected by the surface changes than by the deep-structural change. Also in accordance with the hypotheses, the main effect of expertise and the interaction of expertise with type of rule modification were not significant, $p>.10$.

Insert Figures 3, 4 about here

Discussion

The major purpose of Experiment 2 was to support our claim that expertise and flexibility are strongly linked in a cognitive system. Experiment 1 had demonstrated that different degrees of proceduralization and different knowledge-base structures lead to differences in the ease with which subjects can deal with novel task demands. In Experiment 2, we showed that subject-groups that do not differ on these two variables also do not differ in how flexibly they can use their existing knowledge. The results of Experiment 2 strengthen our claim that the degree of proceduralization and the structure of a knowledge base affect the flexibility of a cognitive system.

Experiment 3

In Experiments 1 and 2, we were primarily concerned with the differential impact of surface-structural and deep-structural rule modifications on expert and nonexpert bridge players and with the mechanisms of interference in the lead-change condition. So far, however, we did not discuss the mechanisms of interference for the two surface-structural rule modifications. Whereas subjects' problems with the lead change can only be explained in terms of conceptual levels of processing, subjects' difficulties with the name and rank-order changes might, either in addition to interference on the conceptual level or solely, be attributed to perceptual phenomena. It might be argued, for instance, that the surface modifications affected all subjects primarily at the level of encoding and only to a minor degree at a conceptual level. Or alternatively, that experts were most affected on a perceptual level whereas nonexperts were most affected on a conceptual level or vice versa. Experiment 3 was designed to test the effects of the name-change and rank-order-change conditions at the level of encoding.

In Experiment 3, players of differing skill levels were shown slides of bridge hands (13 cards displayed in a fanned position) in exactly the same way they usually perceive bridge hands. Slides were visible for only 5 seconds. After slide-offset players were asked to write down as many of the cards as they could remember. They had the option of watching the same slide as often as they wanted to until they had written down all 13 cards of the seen hand. Half of the bridge hands shown were structured in the same way players usually structure their hands; for the other half, features of the visual display were changed so as to correspond to the name-change and rank-order-change conditions used in Experiments 1 and 2. In the name-change condition, the letters on honor-cards were changed to new ones. In the rank-order condition, the familiar rank orders of suits and high cards were changed; consequently hand displays mirrored the new rank orders. Reaction time, number of cards written down per trial, and number of cards correctly identified per trial were recorded.

Method

Subjects

Subjects were the same as in Experiments 1 and 2.

Materials and Procedure

Experimental task. Subjects were seated in front of a computer screen and keyboard. They were shown slides with bridge hands (13 cards per hand) on the wall just above the computer screen in a distance of approximately 1.50 meters. They were told that the bridge hands would stay visible for 5 seconds and that they should try to memorize as many cards as they could. Slide onset and offset were controlled by the computer. As soon as a slide disappeared subjects were prompted to write down as many cards as they could remember using special predesigned keys on the computer keyboard. Subjects had the option to view a slide as often as they wanted. They were instructed to watch a slide as often as they needed to write down all cards of the bridge hand.

Subjects responded to 40 different slides or bridge hands. Half of these slides were shown adopting the commonly used hand structuring (ordered for suits and rank of card within suit); for the other 20 slides the visual display was changed in one of two ways.

Types of display modifications. The two types of rule modifications used in Experiment 3 were the two surface changes used in Experiments 2 and 3. They were a name change and a rank-order change. In the name-change condition, the familiar card and suit names (ace, king, queen, jack, and spade, heart, diamond, club) were replaced by nonfamiliar, nonsense names (rutz, lork, dill, beib for cards, and pular, biref, ramog, kamer for suits, respectively). Thus, the letters appearing in the corners of a card identifying ace (A), king (K), queen (Q), and jack (J) were replaced by R, L, D, and B respectively. In the rank-order-change condition, the familiar card order (ace, king, queen, jack) was changed to king, jack, ace, queen, and the familiar suit rank order (spade, heart, diamond, club) was changed to diamond, heart, club, spade. In this condition, only the structuring of cards within suits was changed in the display. The third condition was a control condition in which subjects saw only regularly structured hands.

Design

Dependent variables were the number of cards subjects wrote down after each 5-sec. glance (Total), the number of cards they identified correctly (Correct), and the average time they needed to write down the cards they remembered (RT). Independent between-subjects variables were level of expertise (nonexpert, expert) and type of display change (name change, rank-order change, control). Slide (1 to 40) was a within-subjects variable. In each expertise group, subjects were randomly assigned to one of the three types of modifications with the restriction that no subject received the type of modification she had experienced in Experiments 1 and 2.

Results

Because some of the subjects were able to identify all 13 cards of a bridge hand after their first 5-sec. glance, data analysis was limited to subjects' first trial on each slide.

Reliabilities of Dependent Variables

Table 23 lists the reliabilities for all dependent variables on both normal-display and modified-display trials. As can be seen, most of the reliabilities are in the high 70's or above. Reliabilities for the modified-display trials are generally lower than those for the normal-display trials. This is particularly true for the subject reliabilities. Overall, however, the reliabilities are quite satisfying.

Insert Table 23 about here

Correlations Among Dependent Variables

Table 24 shows the correlations between all pairs of dependent variables. Of particular interest is the correlation between speed and quality of subjects' responses. As can be seen, reaction time (RT) correlated significantly negatively with the number of cards subjects wrote down (Total) and with the number of cards they remembered correctly (Correct), indicating a speed-accuracy tradeoff betwen speed and quality of subjects' responses. The faster subjects responded the more likely they were to falsely identify cards. The significant correlations among the dependent variables point to the use of multivariate techniques as the appropriate means of data analysis.

Insert Table 24 about here

Basic Statistics

As in Experiments 1 and 2, the within-subjects factor of slide (1 to 40) was broken down into four blocks of 20 slides each. Basic statistics for experts' and novices' performances in the four blocks are shown in Tables 25 and 26. Means on each of the three dependent variables are listed as a function of type of display change and block.

Insert Tables 25 and 26 about here

Base-line Performance

Table 27 shows the means for the two expertise groups on each of the three dependent variables for those trials that were displayed under regular conditions. Because the dependent variables were significantly correlated, a multivariate analysis of variance was performed. For the main effect of expertise, Wilks L=0.764, $F(1,32)=3.09$, $p<.05$. Experts responded faster, wrote down more cards in total, and correctly identified more cards than did novices. In order to decide which of the dependent variables contributed most toward the discrimination of the expertise groups, a follow-up discriminant analysis was conducted. Because there were only two groups to separate, only one non-zero eigenvalue could be obtained, accounting for approximately 8% of the total variance. The structure coefficients of the discriminant function were -0.38, 0.83, and 0.93 for the dependent variables of RT, Total, and Correct, respectively, indicating that experts' encoding superiority reveals itself more clearly in the amount of information they can take in than in the time they need to do so.

Insert Table 27 about here

Changes in Encoding Ability due to Modifications of the Visual Display

Table 28 depicts the mean percent increases or decreases in the three dependent variables (relative to individual base-line performances) for the two expertise groups in all experimental conditions. A two-way multivariate analysis of variance with expertise and type of display modification as independent variables found only the main effect of type of display change to be significant, Wilks L=0.58, $F(2,28)=2.72$, $p<.05$). Both display-change conditions differed from the control condition (Wilks L=0.74, $F(1, 28)=3.08$, $p<.05$ for the name-change condition; Wilks L=0.76, $F(1,28)=2.73$, $p<.07$ for the rank-order-change condition), but not from each other. The main effect of expertise and the interaction of expertise-type of display change were not significant. Thus, experts and nonexperts did not differ in the extent to which they were affected by the display changes in any of the experimental conditions.

Insert Table 28 about here

Discussion

The three major results of Experiment 3 can be summarized as follows: First, replicating Charness's (1979) earlier findings for the domain of bridge, we found that experts encoded meaningful information faster and with fewer errors than did nonexperts. Second, the two types of display changes did not differ in their effects on subjects' encoding abilities. And finally, experts and nonexperts were not differentially affected by the two types of changes of the visual display.

The fact that experts and nonexperts were not differentially affected by the two types of changes has important implications for the discussion of Experiments 1 and 2, suggesting that the effects of the surface-structural rule modifications obtained in these studies might, for both expertise groups, be partly due to the disruption of normally occuring encoding processes. The findings of Experiment 3 do not support the argument that the surface changes affected experts primarily on a perceptual level and nonexperts primarily on a conceptual level of information processing or vice versa.

General Discussion

The three experiments described in this report were conducted to explore the relation between expertise and flexibility. Studies investigating the nature and development of expertise typically focused upon the benefits of mastery in a domain, that is, they addressed the question of why it is that experts perform more quickly and better than nonexperts do. In contrast, we argued that greater expertise does not necessarily result in superior performance in all instances. In fact, we hypothesized that the very same mechanisms that are responsible for experts' sometimes amazing performances on conventional, familiar tasks might work against them when they have to deal with novel task demands. Specifically, we assumed that two variables, the degree of proceduralization and the organization of the knowledge base, can cause two different types of inflexibility, the inability to inhibit existing response patterns, and the inability to modify an existing system of knowledge.

The results of the research presented in this report can be summarized as follows. First, problem solvers who have proceduralized their solution strategies are less flexible than ones who have not. Experts were generally more affected than nonexperts when task demands changed. Second, problem solvers are less flexible when dealing with task modifications that are incompatible with the structure of their knowledge representations than when dealing with changes that are compatible. We found that experts were more affected by deep-structural changes, whereas nonexperts were more affected by surface changes when the two groups structured their knowledge representations differently. When both groups operated on the same knowledge base, they were not differentially affected. Third, all subjects were able to adjust to new task demands. In most conditions, they returned to their original levels of performance quite quickly. And finally, the effects of the two surface-structural rule manipulations might for all subjects be partly due to perceptual, rather than conceptual phenomena.

On a general level, it can be argued that the findings reported in this report are due to negative interference of previously acquired knowledge on the acquisition of new knowledge. The basic form of this interference might be considered a variation of the common A-B, A-C framework, used in studies of associative learning. Bower (1974) pointed out that interference theory need not be restricted to meaningless material or to "factual details." In our work, subjects experienced interference when they were required to reason with partly novel concepts, with information, that is, that was only partially meaningful in terms of prior knowledge.

This particular interpretation is very closely related to Sternberg's (1981, 1982) discussions of "nonentrenched" concepts and Tetewsky and Sternberg's (1986) work on the conceptual and lexical determinants of nonentrenched thinking. Theoretical accounts of negative interference effects in associative learning, as mentioned in the introduction, have emphasized different components of flexibility. Proactive interference (e.g., Underwood, 1957) and list dicrimination theories (e.g., Sternberg & Bower, 1974) describe negative interference effects primarily in terms of subjects' inability to inhibit older, previously established associations, whereas organizational theory (e.g., Tulving, 1966) tries to link them to adaptation difficulties. Our data suggest that the type of interference that occurs might be related to the organization of prior knowledge and to the extent that proceduralized strategies are being used. Apparently, problem solvers with less structured and less proceduralized knowledge systems (nonexperts) have more difficulties inhibiting old responses than they have modifying their knowledge representations. Problem solvers who rely on highly structured and proceduralized knowledge systems (experts) on the other hand, seem to experience most difficulty when modifying prior knowledge.

The experimental findings reported here can also be discussed with regard to recent theoretical developments in the field of intelligence. It has long been a traditional belief that intelligence can be characterized, in part, as the ability to acquire new information. Several theorists have proposed that the ability to deal with novel, and particularly, partially novel information is an integral aspect of intelligence (e.g., Berg & Sternberg 1985; Cattell, 1971; Horn, 1968; Raaheim, 1974; Snow, 1981; Sternberg, 1985). The present findings suggest that this ability can reflect inter-subject differences in domain-specific strategies and categorization of knowledge.

This interpretation also lends itself very easily to a phenomenon that has been reported frequently in the developmental literature: the decline of fluid intelligence with age (Horn, 1968). Our data suggest that this decline might be due, at least in part, to the proceduralization of responses to highly recognizable and familiar situations, and to the organization of existing knowledge bases. In other words, by being more "expert" at performing a range of familiar tasks, older individuals might experience an impairment in performance when a novel task competes with what is already known.

In summary, the research presented here touches upon several important issues within cognitive psychology. We hope to have demonstrated that the costs of expertise can be fruitfully exploited not only to unravel the mysteries of mastery but also to understand basic cognitive principles.

Conceptual and Lexical Determinants of Nonentrenched Thinking[4]

We rely on our knowledge of the world to make inferences and to form expectations. The various schemata and concepts that are stored in our memories permit us to understand the complex relationships that we experience in our everyday lives. Although our knowledge allows us to deal with a broad range of situations, there are limitations to its usefulness—when we encounter something that is unfamiliar, such that it fails to conform to our expectations, we tend to make incorrect inferences. These errors reflect a fundamental aspect of human thought because whenever we experience something new, we have to deal with it in terms of what is familiar and old-fashioned (Oppenheimer, 1956).

The question of why we find it difficult to make valid inferences in unfamiliar situations is implicit in Goodman's (1955) "new riddle of induction." Goodman's philosophical argument, reviewed by Skyrms (1975) and by Sternberg (1982), implies that certain events are quite natural and others are quite unnatural. We expect the former kinds of events, but not the latter kinds of events, to occur. Goodman argued that as a result of our cultural experiences, we develop certain terms to deal with our everyday lives. These "entrenched predicates" cause us to favor certain inferences over others even though the available evidence supports several different conclusions.

Goodman's philosophical ideas about "entrenchment" are related to a psychological phenomenon that was first explored by Barlett (1932). The subjects in Bartlett's experiment were required to remember stories that came from a foreign cultural and social environment. The nature of these stories caused subjects to make a number of systematic errors. Bartlett's work has traditionally been used to support the idea that when we recall a text, we remember a general schema and we use this schema to reconstruct the specific details of the text. However, his work also suggests that once we have a knowledge system that describes how the world is ordered, it is often quite difficult to consider alternative ways of thinking. This interpretation is compatible with the work of Maier (1931) and of Duncker (1945), who showed that prior knowledge can impede problem solving through the mechanism of "functional fixedness."

Goodman's philosophical ideas about entrenched predicates appear to have some interesting psychological implications. The studies cited above suggest that when we try to comprehend information that is "nonentrenched," we experience interference from our prior knowledge. However, it is important to specify the precise meaning of nonentrenchment and the exact nature of the associated interference. In particular, it is necessary to determine whether nonentrenchment is different from unfamiliarity, and it is also necessary to specify the conditions under which prior knowledge will cause interference in processing nonentrenched concepts. The research reported in this article was designed to address these issues by exploring the psychological reality of nonentrenchment.

Several attempts have been made to distinguish between "natural" and "unnatural" concepts and between entrenched and nonentrenched concepts. Rips (1975) dealt with the nature of entrenched concepts by showing that representative (or prototypical) instances of a category facilitate inductive generalizations, compared to less representative (or prototypical) instances of a category. He therefore equated entrenchment with representativeness. Osherson (1978) formally considered the various theoretical issues that might be dealt with in distinguishing between natural and unnatural concepts. For example, he suggested that natural concepts are more likely to figure in law-like generalizations (i.e., they are more "projectible") and that they should facilitate deductive reasoning in comparison to unnatural concepts. Osherson's work was mainly theoretical, rather than empirical, emphasizing the formal constraints on natural concepts. Finally, Keil (1979) has argued that natural concepts are formed out of basic "ontological categories," which refer to the most general categories of things in the world. For example, a natural concept might contain the category of all living things, whereas an unnatural concept might be made of humans and plants, without also including nonhuman animals. Keil's work therefore emphasizes the structural differences between natural and unnatural concepts.

In contrast to these analyses, Sternberg (1982) directly studied the processing implications of natural and unnatural concepts by designing a reasoning task based on Goodman's (1955) new riddle of induction. For example, in one experiment, an unnatural concept was described by the word "grue," which referred to an object that was green in the year 1977 and turned blue in the year 2000. The corresponding natural concept, which did not involve a transformation, was described by the term "green." This term referred to an object that was physically green in the year 1977 and was also green in the year 2000. Sternberg proposed three alternative hypotheses to explain why natural concepts, such as green, might be easier to process than unnatural ones, such as grue: (a) natural concepts belong to the conceptual system with which we grew up and so they are more entrenched in our experience; (b) natural concepts do not carry an expectation of change and so they are "simpler," psychologically, to process; and (c) the effects of simplicity and entrenchment may operate jointly such that natural concepts may be both simpler and more entrenched in our experience. The results of Sternberg's research indicated that both entrenchment and simplicity contributed to ease of information processing, but that the effect of simplicity was greater than the effect of entrenchment.

The present work was undertaken to extend Sternberg's (1982) work by clarifying the nature of nonentrenchment. A reasoning was designed in which the naturalness of a concept and the type of name used to define that concept could be independently varied. A conceptual system was found in which the content could be expressed in four different forms, such that two levels of concepts (natural or unnatural) could be crossed with two levels of names (familiar or novel). The underlying assumption for this design was that concept naturalness and lexical familiarity might be important, singly or in combination, in distinguishing between entrenched and nonentrenched concepts.

In the first experiment to be described, subjects were required to solve reasoning problems in which they had to select among alternative projections about occurrences in the environment that relate to seasonal changes. It is quite natural for the leaves to change color in accordance with the seasons (at least in New England!). However, it is not at all natural for us to think that rocks will change color according to a seasonal pattern. Analogously, seasons can be identified by the names <u>summer</u>, <u>fall</u>, <u>winter</u>, and <u>spring</u>, or they can be given navel names, such as <u>soob</u>, <u>trit</u>, <u>blen</u>, and <u>mave</u>. By using these two sets of concepts and names, the following four situations were constructed: (a) familiar season names describing states of the leaves; (b) novel season names describing states of the leaves; (c) familiar season names describing states of the rocks; and (d) unfamiliar season names describing states of the rocks. In the second experiment, subjects were required to make projections about events in the environment that relate to periods of the day. In this context, it is natural to identify a period of the day by noting the position of the sun relative to the horizon and it is quite unnatural to expect that minerals will change shape as the day progresses. Also, the periods of the day can be identified by the names <u>daytime</u>, <u>dusk</u>, <u>nighttime</u>, and <u>dawn</u>, or they can be given novel names such as <u>trofar</u>, <u>bren</u>, <u>stobe</u>, and <u>kovit</u>. By using these two sets of concepts and names, a set of four situations was constructed that was structurally equivalent to that describing seasons in the first experiment.

Subjects in each of these conditions were given descriptions of the beginning and end of a season (or the beginning and end of a period of the day) and were required to make inferences regarding the events that occurred. The problems were presented in individually as "selection task items." The ease with which subjects made these judgments was measured by both latency and error indices. A model of information processing was also tested for the latency data obtained in each of the four tasks. In addition to requiring solution of selection-task items, these experiments also required subjects to solve sets of reasoning problems, which were of varying degrees of difficulty. Many of these problems were similar to ones drawn from psychometric intelligence tests. The selection task items therefore provided a way to asses the extent to which psychometrically measured intelligence is associated with the ability to reason within new conceptual systems.

Models of Nonentrenchment

These experiments presented an opportunity to compare different structural models for nonentrenched concepts. The potential effects of lexical familiarity and conceptual naturalness on nonentrenchment can be described in terms of five basic models, which are shown in Table 29. By evaluating each of these models, we can determine the extent to which nonentrenchment can be explained in terms of unfamiliarity or interference from prior knowledge. Analysis-of-variance contrast weights are used to represent the patterns of reaction time and error rates that characterize each of these models.

Model 0

In Model 0, the null case, there is no effect for either lexical unfamiliarity or conceptual unnaturalness. This model implies that the construct of nonentrenchment is not psychologically important. In essence, this model represents the null hypothesis.

Model 1

In Model 1, the locus of nonentrenchment can be found entirely in conceptual unnaturalness; according to this formulation, lexical familiarity is not a source of interference.

Model 2

Model 2 shows the complementary situation, in which the locus of nonentrenchment can be found entirely in lexical unfamiliarity; according to this formulation, conceptual naturalness is not a source of interference.

Insert Table 29 about here

Model 3

Model 3 describes the situation in which lexical unfamiliarity and conceptual unnaturalness both contribute to nonentrenchment, such that their effects are additive. This model distinguishes between two levels of nonentrenchment. On one level, represented by "0" in the table, nonentrenchment is characterized by using either familiar names to denote unnatural occurrences of unfamiliar names to denote natural occurrences. On another level, there is a more difficult form of nonentrenchment that involves using unfamiliar names to denote unnatural occurrences. This model therefore involves both unfamiliarity and unnaturalness contributing additively to nonentrenchment. Moreover, if we assume that unnatural concepts are in some sense unfamiliar, this model also indicates that the overall amount of unfamiliarity in a conceptual system defines the extent to which that conceptual system is nonentrenched.

Model 4

Finally, Model 4 describes the situation in which nonentrenchment is defined by an interaction between lexical unfamiliarity and conceptual unnaturalness. According to this model, there are two equivalent types of nonentrenched concepts, one in which familiar names denote unnatural occurrences, and another in which unfamiliar names denote natural occurrences. Nonentrenchment is the result of an interference effect, in which the interference is generated by either a familiar name or a natural concept, but not both simultaneously. This model therefore shows that nonentrenchment is distinct from the overall unfamiliarity of a given conceptual system.

Experiment 1

Method

Subjects. Subjects were 96 Yale undergraduates who participated for course credit, monetary payment, or both. Subjects were randomly assigned to one of four conditions, with 24 subjects in each condition. A separate group of 25 undergraduates taking a developmental psychology course at Yale gave responses to a set of background survey questions.

Materials. The basic materials were selection-task items presented via a tachistoscope and psychometric inductive and deductive ability tests presented in a paper-and-pencil format. In addition to the standardized ability tests, subjects were given a set of problems that have previously been used to study insight (Sternberg & Davidson, 1982).

The selection-task items were modeled on previous problems used by Sternberg (1982). Items were developed in which a common "deep" structure was used to generate four different "surface" structures. In each of the four sets of items, the problems were based on the initial premise that the seasons of the year allow one to predict certain occurrences in nature and that, in turn, these occurrences identify what a given season is. Each problem contained two pieces of information. The first piece of information described a situation at the beginning of a season and the second piece of information provided follow-up data from the end of the same season. Because each of the four variations of the task were similar, only one version will be described in detail. The other variations will be described more briefly.

The premise that served as the model for the other three versions of the task stated that in New Haven, the beginning and end of each season are marked by the fact that the leaves will be either green or brown. In summer, the leaves are green at the beginning and at the end of the season. In fall, the leaves are green at the beginning of the season but are brown at the end. In winter, the leaves are brown at both the beginning and end of the season. And, in spring, the leaves are brown at the beginning of the season but are green at the end. Subjects were required to use this information to solve a series of reasoning problems.

Each problem was presented on one card. Each term of a problem could contain one of two forms of information. The description could be either a picture of the leaves, indicated by a green or brown circle, or the name of a season that represents a decision about what season it is, based on the color of the leaves at the time the observation was made. Information about the leaves at the beginning of the season appeared on the left and information about the leaves at the end of the season appeared on the right. Because each of two descriptions of the leaves (one at the beginning of the season and one at the end) could take either of two physical forms (brown or green) or four verbal forms (an inference based on a season name), there were 6 X 6, or 36, distinct items.

The subject's task was to describe the leaves at the end of the season, based on the information provided in the problem. If the given description for the end of the season was a picture of the leaves, the subject has to indicate the correct name of the season. If the given description for the end of the season was a name, the subject had to indicate the correct color of the leaves. There were always three answer choices, from which the subject had to choose the correct one. These alternatives appeared below the problem stem.

There were four different types of problems. Items either had two season names, a picture followed by a season name, a season name followed by a picture, or two pictures. In the first two types of problems, subjects had to determine the color of the leaves at the end of the season--in the other two problems, subjects has to give the name of the season consistent with the given information. Subjects were alerted to a further complexity in the selection task, which also applies in the real world. At the beginning of a season, it is impossible to distinguish summer from fall or spring from winter, if the only available information is the initial color of the leaves. Also, the names "summer" and "winter" imply that the leaves will remain the same color, whereas the names "spring" and "fall" imply that the leaves will change color by the end of the season. For the problems in which the first term was the name of the season, this name correctly described the color of the leaves at the beginning of the season, but only predicted what color the leaves would be at the end. This prediction might not correspond to the color that was described by the second term. Thus, it was not possible to know for certain the color of the leaves at the end of the season or to know for certain the true season. When the first term of the problem was a picture of the leaves, this complexity did not exist, because a physical description carries no implication regarding the future physical appearance of the leaves.

Although this uncertainty in prediction did not exist for information describing the leaves at the end of the season, there was a related problem associated with the second term. When a season name described the leaves at the end of the season, it could be assumed to provide correct information about both the beginning and the ending color of the leaves, because assessments of the season made late in the season were based on observations of the leaves throughout the entire season. For the problems in which the second term was a name, however, this season name could be "inconsistent" with the starting color of the leaves, as defined in the first term of the problem. For example, if the first term were "summer," this name means that the leaves were green at the beginning of the season and predicts that they would be green at the end. If the second term were "spring," this name means that the leaves were brown at the start of the season and eventually turned green. Because the leaves cannot be both brown and green at the beginning of the season, this problem describes an inconsistent situation; as a result, it was impossible to determine the color of the leaves at the end of the season. The correct answer was thus "inconsistent."

To summarize, physical descriptions, which carried no necessarily correct implication for what the leaves would look like at another time, were always accurate with respect to the appearance of the leaves at the time of the description. However, they might not be accurate with respect to the appearance of the leaves at the end of the season. A complete listing of the 36 distinct problems can be found in the Appendix.

This experiment attempted to assess the extent to which various conceptual systems are more or less "entrenched" by comparing how different problem contents and forms affect reasoning. In the form mentioned above, subjects were required to reason within an entrenched framework. The other three forms of this task varied either the naturalness used, or both. It is expected that leaves will alternate between green and brown as the seasons change. However, it is not at all normal to expect that rocks will change from orange to blue with the passage of seasons. Similarly, the terms summer, fall, winter, and spring carry certain connotations about the seasons they name, but the neologisms soob, trit, blen, and mave do not carry any unequivocal information about the physical world. Because there are two types of concepts (natural and unnatural) and two types of names (familiar and novel), there are 2 X 2 or four possible versions of the season-color information.

In a second condition, subjects were told about the distant country of Latzania, where the leaves change color just as they do in New Haven, but the seasons are called trit, blen, mave, and soob. In a third condition, subjects were told about the planet Kyron, where the seasons are called summer, fall, winter, and spring, but are marked by the fact that rocks change from blue to orange or from orange to blue. In the fourth condition, subjects were told about the planet Kyron, where the seasons are called trit, blen, mave, and soob, and can be distinguished by the fact that the rocks vary from orange to blue, according to a systematic pattern. Each of the four tasks is summarized in Table 30.

Insert Table 30 about here

Using each of the conditions described above, three more sets of 36 problems were generated that were structurally identical to those described for the case of New Haven. The only difference among the four tasks involved the surface structure of the items, as defined by a particular concept--language combination. This manipulation of content made it possible to identify the locus of nonentrenchment.

The ability tests used in the experiment were geometric series completions (Abstract Reasoning) from the Differential Aptitude Test (Bennett, Seashore, & Wesman, 1973), letter and number series completions (Reasoning) from the SRA Primary Mental Abilities, adult level (Thurstone, 1962), and deductive syllogisms (subtest 3) and confirming the validity of conclusions (subtest 9) from the Watson-Glaser (1964) Critical Thinking Appraisal.

The survey requested subjects to rate how common it is, in their experience, for either leaves or rocks to change colors with the seasons. In addition, subjects were given an array with the names "summer," "fall," "winter," and "spring" on one side and the four possible blue-orange or brown-green pairings on the other side. Their task was to match one season name with one of the four physical occurrences. Subjects answered questions about either leaves or rocks, but not both.

Design. The overall design placed subjects within a 2 X 2 between-subjects factorial arrangement. The primary dependent variable was solution latency; a secondary dependent variable was error rate. In item construction, independent variables were the six possible state descriptions [(e.g., in the case of New Haven, summer, fall, winter, spring, G (green circle), and B (brown circle) crossed with the two possible times of occurrence (the beginning and end of a season)]. All subjects saw each of the 36 possible item types three times, with the correct option in a different location each time. The items in each task variant were grouped into three blocks so that a subject had to complete an entire set of 36 items before seeing an item for a subsequent time.

Each subject was randomly assigned to one of the four versions of the selection task. For items in which subjects had to determine the color of the leaves at the end of the season, distractors consisted of an incorrect picture and an "indeterminate" (I) option, to correspond to the possibility that the information in the problem could be describing a self-contradictory and hence, indeterminate situation. For the items that required subjects to determine the name of the season, distractors consisted of two of the three possible word distractors balanced over the three replications of the task for a given experiment. Thus, each possible word distractor appeared equally often across item replications. In the three versions that described the seasons in Latzania or on Kyron, four counterbalanced forms of the items were constructed so that each season name was paired with each of the four physical occurrences (two concepts describing a physical change and two concepts describing constant physical states) only once. For example, in Latzania, the season that would correspond to summer was soob in Form A, trit in Form B, blen in Form C, and mave in Form D. This scheme was followed for each of the other three seasons. This method of counterbalancing was not applied to the New Haven scenario, however, because this scenario did not have any possible alternative forms. Subjects who filled out the survey were given one of two alternative forms of the questions. The forms differed in the order that the season names were listed, so that subjects would not be biased in favor of choosing a particular season name for any one color change.

Apparatus. Selection-task items were administered on an Iconix tachistoscope with attached millisecond timer. Subjects signaled their responses by pressing the button corresponding to the appropriate answer option. Psychometric ability tests, insight problems, and the survey were all administered in written form.

Procedure. Subjects completed an informed consent from and then were given instructions for the appropriate task. The instructions were rather lengthy and required subjects learn how to solve an entirely new set of reasoning problems. Because of the complexity of the task, after each subject finished reading the instructions, the experimenter reviewed the essential elements involved in each of the four types of problems. Then subjects received eight practice items, two of each of the four types of problems described earlier. When needed, extra practice items were provided until subjects were able to give correct responses and demonstrate that they were aware of the different requirements of each problem. Subjects were instructed to solve the items as rapidly as they could under the constraint that they be as accurate as possible. After the practice trials were over, subjects received three randomized blocks of 36 problems, each of a different type. Each of the 36 problem types appeared once in each of the three blocks. Items were drawn on separate 6 X 9-in. cards. Each answer option was correct equally often in each block (12 times per block). The experimenter initiated each trial. The millisecond clock started as soon as the stimulus was illuminated in the tachistoscope. It stopped as soon as the subject pressed one of the three answer buttons. In general, feedback was not provided during selection-task trials, unless subjects made three errors in a row. This feedback was given to ensure that subjects were aware of the various intricacies involved in the different types of problems. Of the 96 subjects who participated in this experiment, 12 were given feedback. The selection task usually took about 1 h. to administer.

The ability tests were administered at a later time in small groups. All tests were timed and subjects were told to complete the tests as quickly and accurately as possible. The various tests were always presented in the following order, under the specified time constraints: (1) letter series (2 1/2 min), (2) deductive syllogisms (6 min), (3) number series (2 1/2 min), (4) confirming the validity of conclusions (4 min), (5) abstract reasoning (12 min), and (6) insight problems (20 min).

Information-Processing Model

A proposed information-processing model for performance in the selection task is shown in Figure 5. Information processing is assumed to be serial with respect to the component processes in the model. Each process is assumed to contribute additively to total solution latency, and each process was assigned a mathematical parameter to represent its duration. This model is similar, but not identical, to the model used in Sternberg (1982).[5]

Insert Figure 5 about here

Results

Basic Statistics

Main experiment. Basic statistics for the groups defined by the 2 X 2 design are provided in Table 31. Each solution time is based on 108 observations per subject, averaged over the 24 subjects in each group. An

Insert Table 31 about here

analysis of variance was done on the reaction time data. The main effect of concept (leaves turning brown and green versus rocks turning orange and blue) was not significant, $F(1,92) = 1.21$, $p > .05$. The main effect of language (common or novel season names) also was not significant, $F(1.92) = 1.87$, $p > .05$. However, the concept X language interaction was highly significant, $F(1,92) = 7.64$, $p < .01$. The analysis of variance on the error data yielded the same pattern of results. The main effects of concept and language failed to reach significance, $F(1,91) = .03$ and $F(1.92) = 1.76$ ($p > .05$ in each case), respectively. The interaction between these two effects was significant, $F(1,92) = 11.00$, $p < .01$. These results show that the items in sets 2 and 3 were the most difficult to reason with because they produced the longest latencies and caused subjects to make the most errors. Thus, as described by Model 4, a nonentrenched concept takes one of two forms: It may require us to identify a natural occurrence with an unfamiliar name, or it may require us to use familiar names to describe occurrences that are unnatural. If a concept uses unfamiliar names to describe unnatural occurrences, it is easy to reason with because it is not subject to interference from prior knowledge.

The correlations between latencies and error rates computed across 24 subjects argue against a speed-accuracy tradeoff--.25, .47, .44, and .25 for sets 1, 2, 3, and 4, respectively. These positive correlations suggest that subjects with longer latencies tended to make more errors. Only the correlations for sets 2 and 3 were significant, $p < .05$, two-tailed.

Correlations between error rates and lantencies across the 36 item types were substantial, in addition to being highly significant--.64, .69, .61, and .45 for sets 1, 2, 3, and 4, respectively. The first three were significant at the .0001 level and the fourth was significant at the .01 level, using two-tailed tests. The fact that subjects made more errors on items with longer latencies may mean that when more processing operations are required, subjects' reasoning capacity is more likely to become overloaded, resulting in incorrect solutions.

Because the two dependent variables were highly correlated, a multivariate analysis of variance was performed. This analysis replicated the results from the two univariate analyses of variance. For the concept main effect, Wilks $\Lambda = .986$, $F(2,91) = .63$, $p > .05$. For the language main effect, Wilks $\Lambda = .972$, $F(2,91) = 1.31$, $p > .05$. Only the interaction was significant: Wilks Λ .870, $F(1,92) = 6.79$, $p < .01$. A follow-up discriminant analysis on the interaction was carried out to determine the extent to which each of the two dependent variables was important in determining this effect. The resulting discriminant function was significant--Wilks $\Lambda = .874$, X^2 $(2, N = 48) = 12.55$, $p < .05$. The standardized discriminant function weights were .469 for latency and .721 for error rates. The latency and error scores were highly correlated with the discriminant function, .75 and .90, respectively. These analyses confirm that the concept X language interaction was not an artifact due to the correlation between dependent measures. Both the latency and error data support the claim that the locus of nonentrenchment lies in combining something familiar with something that is novel. The discriminant analysis also shows that both of the dependent measures provide complementary information in identifying differences between the groups.

Survey. The survey was conducted to provide supporting evidence that within this population, certain occurrences were more natural than others. The frequencies with which subjects described the seasons summer, fall, winter, and spring, as being related to a particular color change, were tabulated. In the first survey, 91% of the responses were in accordance with the predicted relationship; that is, leaves are green throughout summer, in fall they start out green and eventually turn brown, they are brown throughout winter, and in spring they start out brown and eventually turn green, X^2 (1, N = 11) = 189.1, $p < .001$. In the second survey, although subjects agreed that it was not at all common for rocks to change color with the seasons, they inferred that one pattern of changes was more reasonable than any other; 79% of the responses were in line with the idea that in summer rocks should be orange throughout the entire season, in fall they should be orange at the beginning and blue at the end of the season, in winter they should be blue throughout the entire season, and in spring they should be blue at the beginning of the season and orange at the end, X^2 (1, N = 14) = 188.4, $p < .001$.

The results of the first survey were in accordance with the latency and error data described above; the conceptual system in set 1 was rated as being very familiar, and it was also associated with shorter solution latencies and with lower error rates. However, the second survey predicted that one form of set 3 could be interpreted in a meaningful way. Each of sets 2, 3, and 4 was divided into four counterbalanced versions--A, B, C, and D--with six subjects seeing each version. The mean latencies for the subsets in set 3 were 4.83, 5.02, 4.74, and 4.61 s, respectively. Subjects in the survey singled out 3C as being most interpretable. In this version, rocks are orange throughout summer, in fall they start out orange and eventually turn blue, they are blue throughout winter, and in spring they start out blue and eventually turn orange. An analysis of variance on the respective times for the four versions of set 3 indicated that they were not significantly different, however, $F(3,20) = .06$, $p > .05$. Thus, it appears that conceptual system 3C was not easier to reason with. When forced to think about the seasons in terms of rocks turning orange and blue, subjects seem to associate winter with the colder color--blue--and summer with the warmer color--orange. However, this information does not seem to affect latencies in the information-processing task used in this experiment.

Quantitative Tests of the Information-Processing Model

Table 32 shows the results of various statistical evaluations of the information-processing model of selection-task performance. These results are based only on items that were answered correctly. However, the model was also tested on data from the combined correct and incorrect items. There were no noteworthy differences between the two sets of analyses.

Insert Table 32 about here

The model provides a good fit to the latency data and is consistent across all four instantiations of the task. The values of R^2 (proportion of variance accounted for in the latency data by the model) are all in the low .90s. The item reliabilities are all in the mid to high .90s, meaning that the various items are consistent in the processes they are assessing. These model fits are highly significant (with respect to the difference of R^2 from 0), although the residuals were also significant. The model also provides a very good fit to individual subjects' data, with R^2 values for individual subjects ranging from .24 to .91 across the four data sets. The mean individual R^2 values for each of the four groups were .67, .70, .69, and .71, for groups 1, 2, 3, and 4, respectively. The root-mean-square deviation of observed from predicted latencies for individual subjects ranged from .31 to .45, which is quite good given that the grand mean for all items was 4.14 s. The original model developed by Sternberg (1982) provided good fits when applied to the data sets in this experiment. The R^2 values for sets 1, 2, 3, and 4 were .92, .90, .91, and .89, respectively. However, the revised model represents a statistically significant improvement in terms of the parameter estimates. (Other models were tested with less satisfactory results.)

Parameter Estimates for the Information-Processing Model

The regression model provides parameter estimates that describe how long each of the processing steps takes for each version of the task. The parameter estimates (raw regression weights) are presented in Table 33. All five parameters were significantly different from zero across each of the four tasks. They also followed the same pattern of relative difficulty, except for assess initial information (AII) and change in representational system (CRS) in set 3. Thus, the model is measuring processes that are general across several tasks.

Insert Table 33 about here

The estimates in sets 2 and 3 were generally longer than their counterparts in the other two sets. the longer estimates in these two sets correspond to the fact that these two conditions were more difficult to reason with, as shown in the analysis of variance. They also suggest that the effects of nonentrenchment are localized in particular component processes. For example, the justification (JUS) parameter showed the largest increase in weight, suggesting that subjects find it especially difficult to evaluate the validity of their solution when they have to use a familiar name that describes an unnatural occurrence or an unfamiliar name that describes a natural occurrence. The change in physical state (CPS) parameter had the next largest increase that might be attributable to nonentrenchment. This parameter reflects the fact that variable-state concepts involve two colors that occur in a specific sequence, whereas concepts that describe a constancy contain only one color and so subjects do not have to deal with the order of the two colors to know which constant season is being referenced. Thus, in terms of processing effects, nonentrenchment may make it more difficult to store or access the information in these variable-state concepts. The fact that unexpected lack of change (UCVS) and CRS are large in set 3, but not in set 2, suggests that these parameters reflect the difficulty involved in processing the pictures used to represent the unnatural concepts. Similarly, AII is large in set 2 but not in set 3, implying that this parameter is affected mainly by lexical familiarity. Overall, the parameters seem to reflect processes in the reasoning task that are subject to interference from pairing something familiar with something unfamiliar.

Individual Differences Analysis

Scores on the three inductive tests were standardized and added together to obtain an overall score for inductive reasoning ability. The same type of computation was done to combine scores for the two deductive reasoning tests. This form of analysis was done because the composite scores have greater reliability than the individual test scores. These ability measures, in addition to the insight problems, were correlated with each subjects's mean latency, averaged over 108 items. Parameter scores, obtained by fitting the information-processing model to individual subjects' latency data, were also correlated with the ability tests. The correlations are shown in Table 34. (In these analyses, the tests were standardized by treating each of the four groups separately. When the tests were standardized without regard to group membership, the results were virtually identical.)

Insert Table 34 about here

Consider first the correlations with the inductive measure. The correlations of the global latencies with the inductive reasoning composite are the highest. However, they are not consistently high for each of the four groups--group 2 was the exception with a correlation of -.14. Adding in error rate as a second predictor variable did not change this pattern of results--the multiple correlations were -.63, -.15, -.57, and -.40, for groups 1, 2, 3, and 4, respectively. The correlations of parameter scores with the inductive reasoning measure essentially follow this same pattern, with group 2 failing to show significant correlations.

Consider next the correlation of the global task and parameter scores with the deduction measure. The global task shows a pattern of correlations that is analogous to that found with the inductive reasoning measure; however, the correlations with deductive ability are consistently lower. This same basic pattern occurs with the task score correlations.

The insight problems were highly intercorrelated with both the inductive and deductive test scores. The correlations, averaged over the four groups, were .63 with induction and .44 with deduction. This difference in correlations implies that the insight problems are more related to induction--the ability to take the information stated in a problem and to form a conclusion that is not directly related to the given information. In terms of the selection task, the overall response latency for set 3 had, by far, the highest correlation with the insight problems. Moreover, the inductive scores for set 3 had more significant correlations with task parameters than did any of the other three sets. Because the items in set 3 had familiar names mapped onto novel concepts, these results suggest that insight is closely related to the ability to use existing information in the lexicon to describe unfamiliar occurrences in the environment. Insight can therefore be viewed as a specialized form of inducing structure in problem situations.

Although this experiment yielded results that were both reliable and interpretable, it is possible that nonentrenchment is specific to this particular experimental manipulation. Berlin and Kay (1969) have shown that there is something basic about the way our experience with colors is coded in language. Their work suggests that the present set of results may be attributable, at least in part, to the fact that the seasons were defined in terms of colors. It is conceivable that a concept that is described by something other than color may have different properties within this experimental design. Because of this possibility, a second experiment was carried out to show that the language X concept interaction can be demonstrated for other kinds of conceptual systems.

Experiment 2

This experiment was intended to replicate and generalize the findings of the first study. The major differences involved the content of the conceptual system and the words used to describe this content. Therefore, the description of this experiment will focus only on those aspects of the method that differed from that used in the first experiment. The analysis of the results directly parallels the form used in the first experiment.

Method

Subjects. Subjects were 80 Yale undergraduates who participated for course credit, monetary payment, or both. Subjects were randomly assigned to one of the four conditions, with 20 subjects in each condition. It should be noted that 5 subjects were replaced because they had error rates that were near or above the 30% level. A separate group of 24 undergraduates who were participating in other psychology experiments gave responses to a set of background survey questions.

Materials. The selection-task items used in this experiment were structurally identical to those used previously--they differed only in content. The premise that served as the model for the other three versions of the task stated that in New Haven, the beginning and end of each period of the day is marked by the fact that the sun is either above or below the horizon. At dawn, the sun starts out below the horizon but eventually ends up above the horizon. In daytime, the sun is above the horizon at the beginning and at the end of this period. At dusk, the sun starts out above the horizon but eventually ends up below the horizon. And in nighttime, the sun is below the horizon at the beginning and at the end of this period.

To parallel the first experiment, the periods of the day had to be characterized by an "unnatural" occurrence. Because it is not at all normal for the periods of the day to be characterized by minerals that change shape from oval to rectangular, this occurrence served as the unnatural concept. In one variation, subjects were told about the Hafo Indians of Western Canada who use the names stobe, kovit, bren, and trofar to describe the daily pattern of changes in the position of the sun. In another version subjects were told about the planet Kyron, where the periods of the day are called dawn, daytime, dusk, and nighttime, but are marked by the fact that a certain type of mineral changes shape from rectangular to oval or from oval to rectangular. In the final variation, subjects were told about the planet Kyron, where the periods of the day are called stobe, kovit, bren, and trofar, and can be distinguished by the fact that a certain kind of mineral changes shape from rectangular to oval, according to a systematic pattern. Each of the four tasks is summarized in Table 35.

Insert Table 35 about here

The main difference between this set of materials and the earlier one is that the selection-task items were not drawn on cards; instead, they were presented on a computer (CRT) screen. The picture representations of the sun and the horizon were drawn using five hypens and an "o", with the "o" appearing either 0.5 cm above or below the hyphens. The rectangular and oval shapes were 1 X 0.5 cm and they were drawn using a set of graphics characters. The survey was carried out just as it was in the first experiment.

Design. The same design was used in this experiment as in the previous one, including the counterbalancing scheme for the three variations of the New Haven information.

Apparatus. In contrast to the first experiment, the selection task items in this study were presented on a Televideo 950 CRT, driven by a North Star Horizon computer. Subjects signaled their responses by pressing one of three keys on the keyboard. The ability tests were again given in paper-and-pencil format.

Procedure. Because this experiment was run by a computer, there were several procedural differences. Once a subject completed the practice trials, the experimenter no longer played an active role in the experiment. Subjects initiated each trial by pressing the space bar. Following a 1-s delay, a problem appeared on the center of the CRT screen and an internal clock was started. When subjects hit one of the three answer keys, the clock was stopped, the solution time was recorded, and the problem was cleared from the screen. Subjects were allowed to initiate trials at their own rate. The time between trials was not recorded. This procedure allowed subjects the opportunity to ask questions when they were not clear about some aspect of the experiment. Subjects were also notified by the computer when they made three errors in a row; they were then shown the correct answer to these problems in order to ensure that they understood what each problem required. The time needed to complete this part of the experiment varied from 30 min to a full hour. All other aspects of the procedure were the same.

Results

Basic Statistics

Main experiment. Basic statistics for the groups defined by the 2 X 2 design are provided in Table 36. Solution times are again based on 108 observations per subject, averaged over the 20 subjects in each group. The main effect of concept was not significant, $F(1,76) = 2.48$, $p > .05$, and the main effect of language also was not significant, $F(1,76) = 2.99$, $p > .05$. The concept X language interaction did reach significance, $F(1,76) = 3.98$, $p < .05$, replicating the major finding of the first experiment. However, in contrast to the first experiment, the error data did not follow the same pattern as the reaction-time data. The concept main effect was not significant, $F(1,76) = .358$, $p > .05$, but there was a significant main effect for language, $F(1,76) = 4.32$, $p < .05$, and the concept X language interaction failed to reach significance, $F(1,76) = .99$, $p > .05$. An examination of the mean error rates showed that the two conditions that used familiar names (groups 1 and 3) had the highest error rates. It appears that the pattern of results shown in the analysis of variance can be attributed mainly to group 1, which had a different ordinal ranking for errors than it did in the first experiment.

Insert Table 36 about here

Despite the anomaly in the error rates, the patterns in the latency data cannot be dismissed as being a result of a speed-accuracy tradeoff. The correlations between latencies and error rates computed across 20 subjects were positive: .31, .65, .34, and .15 for sets 1, 2, 3, and 4, respectively. Only the correlation for set 2 was significant, $p < .05$, two-tailed. Correlations between error rates and latencies across the 36 item types were .65, .62, .63, and .55 for sets 1, 2, 3, and 4, respectively. All of these correlations were significant at the .01 level or better, using two-tailed tests.

Because the dependent variables were once again correlated, a multivariate analysis of variance was performed. This analysis showed that the two main effects, as well as the interaction, all failed to reach significance. This deviation from the results obtained in the first experiment can probably be attributed to the aberration in the rank order of error rates, noted above.

In summary, although these results do not exactly parallel those obtained in the first experiment, they do support the basic definition of nonentrenchment, as described by Model 4.[6]

Survey. In the first survey, 89% of the responses were in accordance with the predicted relationship; that is, at dawn, the sun starts out below the horizon and eventually moves above the horizon; in daytime, the sun is above the horizon during the entire period; at dusk, the sun starts out above the horizon and eventually moves below the horizon; and in nighttime, the sun is always below the horizon, X^2 (1, N = 9) = 161.7, $p < .001$. These results are consistent with the latency data in showing that this system describes our entrenched way of thinking. Once again, although subjects agreed that it was not common for the periods of the day to be associated with changes in the shape of certain minerals, 67% of the respondents agreed that one pattern of changes was more reasonable than any other; that is, at dawn, minerals start out oval and eventually become rectangular; in daytime, they are rectangular throughout; at dusk, they are rectangular at the beginning and oval at the end; and in nighttime, they are oval throughout, X^2 (1, N = 15) = 145.5, $p < .001$. Of the four counterbalanced versions of set 3, A, B, C, and D, subjects in the survey identified 3A as being the most meaningful. The mean latencies for the four subsets of set 3 were 6.07, 7.06, 6.51, and 8.04, respectively. An analysis of variance on these four forms indicated that they were not significantly different, $F(3,16) = .51$, $p > .05$. As in the first experiment, it appears that subjects in the survey may have associated the shapes with the relative lightness or darkness of the various periods, but this comparison did not affect performance in the reasoning task.

Quantitative Tests of the Information-Processing Model

Table 37 shows the results from applying the information-processing model to this data set. Once again, these data are based only on items that were answered correctly, although there were no noteworthy differences for the latencies that included trials with errors.

Insert Table 37 about here

In addition to these group statistics, the model was able to account for individual subjects' latency data, with R^2 values for each of the four groups were .70, .64, .64, and .70, respectively. The regression F for group 4 is lower than it is for the other groups, although it is still highly significant. Also, the residuals for group 1 were not significantly correlated. Despite these differences, the various statistics shown in Table 39 indicate that the proposed model also provides a very good fit for the data in this experiment.

Parameter Estimates for the Information-Processing Model

The parameter estimates are presented in Table 38. All the parameter estimates were significantly different from zero across the four groups, showing the generality of the processes being measured. Once again, the CRS parameter was large only in set 3 and not in set 2. However, in general the longest estimates are not consistently found in sets 2 and 3, as in the first experiment, and the five estimates do not follow a consistent order in terms of their relative size. These results may stem from the fact that the subject reliabilities were lower than they were in the first experiment: .82, .73, .72, and .77, for sets 1, 2, 3, and 4, respectively, compared to .84, .85, .86, and .86, in the first experiment. Hence, the first experiment may provide more interpretable information regarding the various processes that are influenced by nonentrenchment because these parameter estimates are based on more stable data.

Insert Table 38 about here

Individual Differences Analysis

The test scores were again combined to form composite scores for inductive and deductive reasoning ability. Correlations of psychometric test scores with the mean latencies and the parameter scores are shown in Table 39.

Insert Table 39 about here

As in the other aspects of this study, the correlations are similar to those in the first experiment, but there are some clear differences. The highest correlations are found in groups 2 and 3, the two versions using nonentrenched concepts. Although the correlations of reaction time with the psychometric test scores are interpretable in terms of a relationship between intelligence and the ability to reason with nonentrenched concepts, they do not correspond to the results of the first experiment. Perhaps the most important regularity to be found in these experiments is the fact that set 3 had the longest mean reaction time, as well as many of the highest correlations. These results suggest that using familiar names to denote unnatural occurrences may be the more interesting form of nonentrenchment; this was the most difficult conceptual system to reason in and it was more highly related to psychometric measures of reasoning. Because this work represents an early attempt to explore nonentrenchment, this issue may bear further research.

General Discussion

The experiments described in this report were carried out to expand on Sternberg's (1982) initial investigations of nonentrenchment. The present work examined how the variables of lexical familiarity and conceptual naturalness function in making a concept more or less entrenched in our everyday way of thinking. The results showed that these variables do not have independent effects on reasoning. As described by Model 4, the locus of nonentrenchment lies in mapping unnatural occurrences onto familiar names or natural occurrences onto novel names. Thus, information that is entrenched is not simply familiar, and information that is nonentrenched is not simply unfamiliar.

Because these experiments used a specially designed reasoning task, it was important to assess the internal and external validity of this task. An information-processing model for the reasoning task was internally validated by showing that it provides a good account of latency data in two experiments, and was externally validated by showing that parameter estimates are correlated with psychometrically derived reasoning scores. The model also provides some explanatory information concerning the various processes that might be affected by nonentrenchment.

This research used Nelson Goodman's (1955) philosophical ideas as a basis for exploring the psychological reality and nature of nonentrenchment. Although the empirical analysis of this problem was straightforward, the interpretation of these results raises several important issues that are concerned with how this research relates to other issues in cognitive psychology.

First, our analysis of the entrenchment construct is based on the finding that highly overlearned names or concepts can produce interference when paired with unfamiliar names or concepts. The basic form of this interference is, of course, not new—it is a variation on the A-B, A-C paradigm from associative learning. The nature of the negative transfer effect within this paradigm has been analyzed in terms of several different frameworks, such as proactive interference (Underwood, 1957), organizational theory (Tulving, 1966), and list discrimination (Sternberg & Bower, 1974). We would like to argue that nonentrenched thinking represents a particularly interesting form of negative transfer. Bower (1974) pointed out that nothing in interference theory restricts its applicability to meaningless materials or to "factual details." He demonstrated that basic interference effects can be produced using simple prose passages. In our work, subjects experienced interference because they were trying to reason with new kinds of information that were only partially meaningful in terms of their prior knowledge. The nonentrenched conceptual systems were designed to be "incorrect" or implausible for the subjects who participated in these experiments; however, these same conceptual systems were assumed to be entrenched for other subjects from a different background. (See Keil, 1979, for a discussion of implausibility and the related phenomenon of anomaly.) The distinction between entrenched and nonentrenched information may provide a different way to think about various problems in learning and memory, although it will remain for future research to support this possibility.

Second, we need to consider why a conceptual system that uses unfamiliar names to denote unnatural occurrences is no more difficult to reason with than an entrenched conceptual system. Although we have explained the difficulty associated with nonentrenched thinking in terms of negative transfer from prior knowledge, it is possible that there is also a facilitation effect that is operating within this paradigm. Bower (1974) showed that the conceptual macrostructure of a text can be a source of retroactive facilitation, whether the underlying details of the text are changed or kept the same. However, other research (Bower, Black, & Turner, 1979, Experiment 5; Kintsch, Mandel, & Kozminsky, 1977) has shown that a conceptual macrostructure can be a source of proactive interference when subjects are required to remember jumbled texts. In our studies of nonentrenchment, we used four conceptual systems that had different "surface" structures (contents) generated from a common "deep" structure (conceptual macrostructure). Hence, the common conceptual structure may have been a source of proactive facilitation. It will remain for future research to demonstrate the possibility that there are also forms of nonentrenched thinking in which a familiar conceptual structure can be a source of negative transfer and proactive interference.

Third, the present experiments were specifically designed to study the nature of nonentrenchment. However, they implicitly dealt with another problem--the relationship between language and thought. The language-thought problem deals, in part, with the following issue: Every language must be flexible enough to express the diverse experiences of a culture; yet, a specific language appears to shape or bias the way the members of a culture think about their experiences. (See Clark and Clark, 1977, for a summary of various approaches to this problem.) Our research relates to this problem in that it demonstrates a situation in which the effects of linguistic and conceptual factors are not independent. Although the potential relationship between these two lines of work is interesting, it is important to limit the scope of this observation. Nonentrenchment was operationalized using two very narrowly defined variables. Lexical familiarity does not begin to capture the structural and semantic complexity of language. And most concepts cannot be described using only two dichotomous states. Hence, the results reported here should be viewed as providing a basis for further research on nonentrenchment and its relationship to the language-thought problem.

Fourth, Goodman's (1955) ideas were originally intended to provide philosophers with a different perspective on the nature of inductive logic. Our empirical analysis of nonentrenchment may therefore have implications for the way psychologists have studied the way we make inductive judgments. Much of the psychological research in this area has attempted to describe the heuristics and biases that we use in making inductive judgments (Kahneman, Slovic, & Tversky, 1983). Our research may relate to this literature in that when we are required to make judgments under uncertain conditions, we are essentially reasoning in nonentrenched conceptual systems. Ideas about base-rate probabilities, sample sizes, and covariation, for example, are nonentrenched; our intuitive theories about these concepts are not easily modified and they interfere with our attempts to make judgments that would be consistent with these concepts. Hence, it may be useful to explore how the construct of nonentrenchment relates to specific heuristics that people appear to use.

Finally, the experimental findings on nonentrenchment are in accordance with the traditionally held belief that intelligence is characterized by the ability to acquire new knowledge. We have drawn our conclusions by assessing individual differences in the ability to solve reasoning problems, according to a componential framework (Sternberg, 1977). Raaheim (1974) has formally discussed many of these basic issues in his analysis of "problem situations." According to his theory, intelligence is most important to problem solving when the problem situation possesses a medium amount of familiarity. He argued that if a problem is very familiar, then it is mastered quite easily; if it is very unusual, then it is not a good indicator of intelligence because a chance, exploratory approach often provides the best or even only way to obtain a solution. The present data on nonentrenched concepts are in accordance with this analysis, in addition to other related theories put forth by Berg and Sternberg (1985), Cattell (1971), Horn (1968), Kaufman and Kaufman (1983), Snow (1981), and Sternberg (1985). However, it is important to note that Schank (1980) has argued that conventional problem solving is an inappropriate way to understand intelligence because it is dependent on script-controlled behavior. Instead, he has stated that intelligence is characterized by "...the ability to make generalizations from completely new situations that are useful for future needs...." Schank has also asserted that prior cultural information is not helpful in understanding intelligence because any test that relies on this existing knowledge will simply be measuring one's prior training. However, his analysis is not sensitive to the fact that cultural knowledge has to be acquired, and that this learning occurs over a number of years. Our work on nonentrenchment indicates that Schank's statement is too extreme--it can actually be relatively easy to reason in entirely new conceptual systems. As noted by Sternberg (1981), intelligence is, at least in part, the ability to apply existing knowledge structures to new conceptual systems.

In summary, research on nonentrenchment is potentially relevant to several different problems within cognitive psychology. Just as Wittgenstein's (1953) work influenced the way psychologists studied concepts, Goodman's (1955) ideas about entrenched predicates may influence the way we think about various problems in learning and memory, induction, and intelligence.

Appendix A

Basic bridge rules

Following are the official rules of bridge quoted from Borland International's handbook of Turbo Gameworks (1985a).

Bridge is a card game for four players, two pairs of partners. The partners sit across the table from one another, with an opponent on each side. For convenience, the players might be referred to as "North", "East", "South", and "West" to correspond with their geographical positions around the table. Thus North and South are partners, East and West are partners, and around the table clockwise are North, East, South, and West.

Each player is dealt 13 cards. The object of the game is for one partnership to contract to win a certain number of tricks, while the opposing pair attempts to prevent this. A trick consists of four cards, one played by each player. Since each player has 13 cards, there are 13 tricks available.

The play to a trick is as follows: one player plays any card, and the other three players each play a card in turn, going clockwise around the table. A player must "follow suit," that is, play a card of the same suit as the first card of the trick if she is able to do so--if not, she may play any card. The winner of the trick is the player who plays the highest card of the suit led (ace is high, deuce is low), unless there is a trump suit (a special suit determined by the bidding) in which case the highest trump played wins the trick. The person who wins a trick plays first to the next trick, and play continues until all 13 tricks have been played.

The determination of the trump suit and which pair contracts for how many tricks is done by an auction--the highest bidder gets the contract. The rank order of the suits is clubs (lowest), diamonds, hearts, spades, and notrump. The smallest number of tricks one can contract for is seven. Six is added to the bid, and that is the number of tricks which the bidder is contracting to win. If two bids contract for the same number of tricks, the one with the higher ranking suit is the higher bid, but a bid for more tricks always outranks a bid for fewer tricks.

The auction proceeds as follows: starting with the dealer, and going clockwise, each player has the option of either passing or making a bid higher than the last bid. In addition, if an opponent has made the last bid one has the option of doubling, and if an opponent has doubled one's bid, of redoubling. These doubles and redoubles increase both the rewards for making the contract and the penalties for losing the contract. The bidding continues until there are three consecutive passes. One is permitted to make a legitimate bid even if one has passed earlier in the auction.

The primary goal in the bidding is to choose the final contract such that the partnership's best suit is trumps, and to contract for game if the partnership can win enough tricks. Since high cards win tricks, evaluation of the strength of a hand depends primarily upon the high card content. The evaluation technique used by almost all players today is as follows:

ace	= 4
king	= 3
queen	= 2
jack	= 1.

There are a total of 40 high card points. Players usually want a hand to be at least one king above average, 13 or more points, to make the first bid, although this requirement can be relaxed when one has a long suit (5 or 6 cards), since long suits produce extra tricks in the play.

When the bidding has concluded, the pair that made the highest bid is the declaring side, and the player of that pair who first named the suit of the final contract is the declarer.

The play proceeds as follows: the player to the left of the declarer leads to the first trick. After this lead, declarer's partner (called the dummy) puts his or her cards face up on the table. The dummy has no further say in the proceedings--the declarer plays both dummy's cards and her own. Play to the first trick continues with declarer playing a card from dummy, then the player to declarer's right playing to the trick, and finally declarer playing the trick. The winner of the trick leads to the next trick, and play continues until all 13 tricks have been completed. If declarer has won at least the number of tricks contracted for, she has made the contract; if not, she has been defeated.

If the declaring side fulfills their contract, they score points, determined by the number of tricks taken, the contract, and the trump suit. If the contract is defeated, the defenders score points.

Appendix B

Skill Level Category System

Category	Brief description
1 Beginner	no experience
2 Novice	some experience, little ability.
3 Advanced Novice	experience, some ability, maybe some conventions.
4 Post-Novice	20+ points, can occasionally place in open games.
5 Average	normal ability level, can play a cohesive system, not very experienced.
6 Bad life master	usually misapplies knowledge, but can play by rote.
7 Above average	can play well, has some theory, often fast-rising, mistakes due to inexperience.
8 Average life master	experienced, can often hold their own with top players.
9 Strong	usually does well in tournaments.
10 Expert	a string of tournament successes, maybe plays professionally.

Appendix C

Cover Story for Game Modification 1--Change of Card and Suit Names

I hope you enjoyed the games you played so far. Right now things are becoming a bit more complicated.

Imagine the following situation: You are traveling in a foreign country and are not able to understand a word of the language people are speaking. One evening, you are invited to join a group of foreigners in your hotel to play a card game. Very soon you realize that the game people are trying to explain to you is exactly the game of bridge you know from home--only with different names for cards and suits.

CARDS:
JACK is now BEIB
QUEEN is now DILL
KING is now LORK
ACE is now RUTZ

SUITS:
CLUB is now KAMER
DIAMOND is now RAMOG
HEART is now BIREF
SPADE is now PULAR.

Everything else appears to be exactly the same as in the game as you know it.

Please try to cope with that situation in the following games and play as you did during the previous three games. There is no need to learn the new expressions by heart. The computer will prompt you with the new expressions when it is your turn to make a move.
If you want to see the new expressions again please hit <A>.
When you are ready to start the next game hit SPACE.

Appendix D

A Complete Listing of the 36 Different Problems for Experiment 1

	Item		Answer
1A	Summer	Summer	G
1B	Winter	Winter	B
2A	Summer	Winter	I
2B	Winter	Summer	I
3A	Summer	Fall	B
3B	Winter	Spring	G
4A	Summer	Spring	I
4B	Winter	Fall	I
5A	Fall	Summer	G
5B	Spring	Winter	B
6A	Fall	Winter	I
6B	Spring	Summer	I
7A	Fall	Fall	B
7B	Spring	Spring	G
8A	Fall	Spring	I
8B	Spring	Fall	I
9A	G	Summer	G
9B	B	Winter	B
10A	G	Winter	I
10B	B	Summer	I
11A	G	Fall	B
11B	B	Spring	G
12A	G	Spring	I
12B	B	Fall	I
13A	Summer	G	Summer
13B	Winter	B	Winter
14A	Winter	G	Spring
14B	Summer	B	Fall
15A	Fall	G	Summer
15B	Spring	B	Winter
16A	Spring	G	Spring
16B	Fall	B	Fall
17A	G	G	Summer
17B	B	B	Winter
18A	G	B	Fall
18B	B	G	Spring

Note. The items for Experiment 2 can be generated by replacing "summer," "fall," "winter," and "spring," with the names "daytime," "dusk," nighttime," and "dawn," and by substituting "A" (above the horizon) and "B" (below the horizon) in place of "G" (green) and "B" (brown).

References

Adelson, B. (1981). Problem solving and the development of abstract categories in programming languages. Memory and Cognition, 9, 422-433.

Adelson, B. (1984). When novices surpass experts: The difficulty of a task may increase with expertise. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10, 483-495.

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369-406.

Asimov, I. (1979). The singing bell. In I. Asimov, M. H. Greenberg, & C. G. Waugh (Eds.), The thirteen crimes of science fiction (pp. 221-237). Garden City, NY: Doubleday.

Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge: Cambridge University Press.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1973). Differential Aptitude Tests. New York: Psychological Corporation.

Berg, C. A., & Sternberg, R. J. (1985). Response to novelty: Continuity versus discontinuity in the developmental course of intelligence. In H. W. Reese & L. P. Lipsitt (Eds.), Advances in child development and behavior (Vol. 19, pp. 2-47). New York: Academic Press.

Berlin, B., & Kay, P. (1969). Basic color terms: Their universality and evolution. Berkeley, CA: University of California Press.

Borland. (1985a). Owner's handbook of turbo gameworks. Scottsvalley, CA: Borland International, Inc.

Borland. (1985b). Turbo bridge. Scottsvalley, CA: Borland International, Inc.

Bower, G. H. (1974). Selective facilitation and interference in retention of prose. Journal of Educational Psychology, 66, 1-8.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. Cognitive Psychology, 11, 177-220.

Cattell, R. B. (1971). Abilities: Their structure, growth and action. Boston, MA: Houghton-Mifflin.

Cattell, R. B., & Cattell, A. K. (1963). Test of g: Culture fair, Scale 3). Champaign, IL Institute for Personality and Ability Testing.

Charness, N. (1979). Components of skill in bridge. Canadian Journal of Psychology, 33, 1-16.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. Cognitive Psychology, 4, 55-81.

Cheng, P. W. (1985). Restructuring versus automaticity: Alternative accounts of skill acquisition. Psychological Review, 92, 414-423.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Clark, H. H., & Clark, E. V. (1977). Psychology and language. New York: Harcourt, Brace, Jovanovich.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change"--or should we? Psychological Bulletin, 74, 68-80.

Davidson, J. E., & Sternberg, R. J. (1984). The role of insight in intellectual giftedness. Gifted Child Quarterly, 28, 58-64.

DeGroot, A. D. (1965). Thought and choice in chess. Paris: Mouton.

Dreman, D. (1982). The new contrarian investment strategy. New York: Random House.

Duncker, K. (1945). On problem solving. Psychological Monographs, 58 (Whole No. 270).

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). French Kit of Reference Tests for Cognitive Factors. Princeton, NJ: Educational Testing Service.

Engle, R. W., & Bukstel, L. (1978). Memory processes among bridge players of differing expertise. American Journal of Psychology, 91, 673-689.

French, J. W., Ekstrom, R. B., & Price, I. (1963). Kit of reference tests for cognitive factors. Princeton, NJ: Educational Testing Service.

Goodman, N. (1955). Fact, fiction, and forecast. Cambridge, MA: Harvard University Press.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. Psychological Review, 75, 242-259.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized ability intelligences. Journal of Educational Psychology, 57, 253-270.

Hunt, E. B., Frost, N., & Lunneborg, C. (1973). Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), The psychology of learning and motivation (Vol. 7). New York: Academic Press.

Hunt, E. B., Lunneborg, C., & Lewis, J. (1975). What does it mean to be high verbal? Cognitive Psychology, 7, 194-227.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1983). Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.

Kalmar, D., & Sternberg, R. J. (1985). When will the milk spoil? Everyday induction in daily life. Boston: Paper presented at the 26th annual meeting of the Psychonomic Society.

Kaufman, A. S., & Kaufman, N. L. (1983). Kaufman assessment battery for children (K-ABC). Circle Pines, MN: American Guidance Service.

Kay, D. S., & Black, J. B. (1985, August). The evolution of knowledge representations with increasing expertise in using a system. Proceedings of the Seventh Annual Conference of the Cognitive Science Society. Irvine, CA.

Keil, F. C. (1979). Semantic and conceptual development: An ontological perspective. Cambridge, MA: Harvard University Press.

Kintsch, W., Mandel, T. S., & Kozminsky, E. (1977). Summarizing scrambled stories. Memory and Cognition, 5, 547-552.

Kuhn, T. S. (1970). The structure of scientific revolutions (2nd ed.). Chicago, IL: University of Chicago Press.

Maier, N. R. F. (1931). Reasoning in humans. II. The solution of a problem and its appearance in consciousness. Journal of Comparative Psychology, 12, 181-194.

Marr, D. B., & Sternberg, R. J. (1986). Analogical reasoning with novel concepts: Differential attention of intellectually gifted and nongifted children to relevant and irrelevant novel stimuli. Cognitive Development, 1, 53-72.

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. Cognitive Psychology, 12, 252-284.

Osherson, D. N. (1978). Three conditions on concept naturalness. Cognition, 6, 263-289.

Piaget, J. (1972). The psychology of intelligence. Totowa, NJ: Littlefield, Adams.

Raaheim, K. (1974). Problem solving and intelligence. Oslo: Universitetsforlaget.

Raaheim, K. (1984). Why intelligence is not enough. Bergen, Norway: Sigma Forlag.

Reitman, J. S. (1976). Skilled perception in Go: Deducing memory structures from inter-response times. Cognitive Psychology, 8, 336-356.

Rips, L. J. (1975). Inductive judgments about natural categories. Journal of Verbal Learning and Verbal Behavior, 14, 665-681.

Schank, R. (1980). How much intelligence is there in artificial intelligence? Intelligence, 4, 1-14.

Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 484-494.

Skyrms, B. (1975). Choice and chance: An introduction to inductive logic (2nd ed.). Encino, CA: Dickenson.

Snow, R. E. (1980). Aptitude processes. In R. E. Snow, P.-A. Federico, W. E. Montague (Eds.), Aptitude, learning, and instruction (Vol. 1): Cognitive process analyses of aptitude (pp. 27-63). Hillsdale, NJ: Erlbaum.

Snow, R. E. (1981). Toward a theory of aptitude for learning. I. Fluid and crystallized abilities and their correlates. In M. Friedman, J. P. Das, & N. O'Conner (Eds.), Intelligence and learning (pp. 345-362). New York: Plenum.

Sternberg, R. J. (1977). Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1981). Intelligence and nonentrenchment. Journal of Educational Psychology, 73, 1-16.

Sternberg, R. J. (1982). Natural, unnatural, and supernatural concepts. Cognitive Psychology, 14, 451-458.

Sternberg, R. J. (1985). Beyond IQ: A triarchic theory of human intelligence. New York: Cambridge University Press.

Sternberg, R. J., & Bower, G. H. (1974). Transfer in part-whole and whole-part free recall: A comparative evaluation of theories. Journal of Verbal Learning and Verbal Behavior, 13, 1-26.

Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M. (1981). People's conceptions of intelligence. Journal of Personality and Social Psychology, 41, 37-55.

Sternberg, R. J., & Davidson, J. E. (1982). The mind of the puzzler. Psychology Today, 16, June, 37-44.

Sternberg, R. J., & Davidson, J. E. (1983). Insight in the gifted. Educational Psychologist, 18, 51-57.

Sternberg, R. J., & Gardner, M. K. (1983). Unities in inductive reasoning. Journal of Experimental Psychology: General, 112, 80-116.

Sternberg, R. J., & Gastel, J. (in press). If dancers ate their shoes: Inductive reasoning with factual and counterfactual premise. Memory and Cognition.

Sullivan, E. T., Clark, W. W., & Tiegs, E. W. (1963). California Test of Mental Maturity. California Test Bureau.

Tetewsky, S. J., & Sternberg, R. J. (1986). Conceptual and lexical determinants of nonentrenched thinking. Journal of Memory and Language, 25, 202-225.

Thurstone, T. G. (1962). Primary mental abilities. Chicago, IL: Science Research Associates.

Tulving, E. (1966). Subjective organization and effects of repetition in multi-trial free-recall learning. Journal of Verbal Learning and Verbal Behavior, 5, 193-197.

Underwood, B. J. (1957). Interference and forgetting. Psychological Review, 64, 49-60.

Watson, G., & Glaser, E. M. (1964). Critical thinking appraisal (Form YM). New York: Harcourt, Brace & World.

Wittgenstein, L. (1953). Philosophical investigations. New York: Macmillan Co.

Footnotes

[1]This research was done in collaboration with Joyce Gastel.

[2]This research was done in collaboration with Joyce Gastel.

[3]This research was done in collaboration with Peter Frensch.

[4]This research was done in collaboration with Sheldon Tetewsky.

[5]The revised model used five rather than seven parameters to predict latencies on 18 distinct items. To do this, a scheme was devised in which three of the parameters, VWP1--"variable state word in position 1," WP--"number of words in premise," and WA--"number of words in answer," were consolidated into one parameter, ALL--"assess initial information in the first premise." An entirely new parameter, JUS--"justification," was added. This parameter dealt with the fact that for "inconsistent" problems that had a variable-state word in the second position, subjects would have to JUSTIFY that they used the correct variable-state season to solve the problem. Also, the nonentrenched parameter was not used because in the present experiments, nonentrenchment was defined by the variables of conceptual unnaturalness and lexical unfamiliarity, within the analysis of variance design.

[6]One possible explanation for the various differences may be that the second experiment allowed subjects to do the task without interacting with the experimenter. Of the two experiments, the second one had higher mean error rates; in addition, several subjects had to be replaced, due to excessive error rates--this problem did not occur in the first experiment.

Table 1

Representative Verification Items from the Experiment

	Keyed Response: Familiar Presupposition	Keyed Response: Counterfactual Presupposition
1. Familiar presupposition: Trees need water. Counterfactual presupposition: Trees eat people.		
a. Librarians eat maple trees.	False	False
b. Cherry-picking requires great bravery.	False	True
c. Trees are carnivorous.	False	True
d. Trees have branches.	True	True
e. Trees are harmless.	True	False
f. Lumberjacks chop down trees.	True	True
2. Familiar presupposition: Cats are furry. Counterfactual presupposition: Cats are strongly magnetized.		
a. Cats "stick" to refrigerators.	False	True
b. Cats attract paperclips.	False	True
c. Cats have sharp claws.	True	True
d. Cats eat iron filings.	False	False
e. Cats like eating fish.	True	True
f. Catnip is metallic.	False	False
g. Suspended cats tend to face east-west.	False	False
3. Familiar presupposition: Kites fly in the air. Counterfactual presupposition: Kites run on gasoline.		
a. Kites emit exhaust.	False	True
b. Kites need fuel.	False	True
c. Kites have tails.	True	True
d. Kites need wind.	True	False
e. Kites have four wheels.	False	False
f. Kites are faster than airplanes.	False	False
g. Kites can explode when they crash.	False	True
h. Kites are sold in stores.	True	True

Table 2

Basic Statistics

	Response Time		Error Rate	
Non-Novel Presupposition	Mean	STD	Mean	STD
True Statements	2.20	.57	.02	.03
False Statements	2.36	.72	.04	.04
Novel Presupposition				
True Statements	2.57	.73	.06	.05
False Statements	2.91	1.02	.13	.09

Note: Response time is measured in seconds.

Table 3

Analysis of Variance for Response Times

	df	Anova MS	F
Presupposition (Novel, Non-novel)	1	852906	80.70***
Subject X Presupposition	49	10369	
Change in Response	1	434	.05
Subject X Change	49	9465	
Answer (True or False)	1	130572	17.29***
Subject X Answer	49	7552	
Presupposition X Change	1	3688	.92
Subject X Presupposition X Change	49	4020	
Presupposition X Answer	1	1189	.30
Subject X Presupposition X Answer	49	2516	
Change X Answer	1	14346	2.10
Subject X Change X Answer	49	6816	
Presupposition X Change X Answer	1	38940	25.73***
Subject X Presupposition X Change X Answer	49	1514	

Note: Response time was adjusted for the number of words in the statement.

*** $p < .01$

Table 4

Correlations of Response Time and Error Rate with Ability Scores

Cue		Letter Sets	CIMM Syllogisms	Cattell Culture Fair	Crossing Out As
Non-Novel	RT	-.46***	-.66***	-.32*	-.04
	ER	-.18	.02	.04	-.15
Novel	RT	-.52***	-.68***	-.32*	-.02
	ER	-.16	-.24	-.11	-.15

* $p < .05$

*** $p < .001$

Table 5

Taxonomy of Induction Items in Four Phases of Induction Research

	Similarity to IQ-Test Items: IQ-Test-Like	Un-IQ-Test-Like
Familiar	Analogies Classifications Series Completions	Everyday Inductions
Unfamiliar	Novel Analogies Novel Classifications Novel Series Completions	Conceptual Projections Insight Problems

Table 6

Sample Stimulus Task Items

Analogies

Familiar Relevant
Chalk is used for writing.
INK : PAPER :: CHALK : (a) WORD, *(b) BLACKBOARD, (c) ERASER, (d) CLASSROOM

Familiar Irrelevant
Beds have sheets.
BUILDING : DOME :: BED : *(a) CANOPY, (b) PILLOW, (c) HOTEL, (d) ROOM

Novel Relevant
The hand is the organ of hearing.
EYE : BLINDING :: HAND : (a) SMOOTH, (b) TOUCHING, *(c) DEAFENING, (d) WORN
(Uncued condition: DEAFENING is replaced by NUMBING)

Novel Irrelevant
Sparrows play hopscotch.
TROUT : SCALY :: SPARROW : (a) GRACEFUL, *(b) FEATHERY, (c) SMALL, (d) ENERGETIC

Classifications

Familiar Relevant
Many fish are edible.
TUNA, SARDINE, SOLE, FLOUNDER, (a) BAIT, (b) TADPOLE, (c) SEAFOOD, *(d) SALMON

Familiar Irrelevant
Soup is mainly water.
BOUILLON, EGG DROP, BROTH, MINESTRONE, (a) SALAD, *(b) CHOWDER, (c) LIQUID, (d) CRACKERS

Novel Relevant
Flowers grow underground.
TULIP, DAFFODIL, ROSE, DAISY, (a) DIRT, (b) PETAL, *(c) ONION, (d) FLOWERPOT
(Uncued condition: ONION is replaced by BUTTERCUP)

Novel Irrelevant
Water boils at room temperature.
FOG, STEAM, VAPOR, CLOUD, (a) PUDDLE, (b) ICE, *(c) MIST, (d) RAIN

Table 6 (continued)

Series Completions

Familiar Relevant

Some people are heavier than others.
SKINNY : SLIM : AVERAGE : PLUMP : (a) HUNGRY, (b) THIN, *(c) FAT, (d) ATHLETIC

Familiar Irrelevant

Feet are bigger than hands.
THUMB : INDEX FINGER : MIDDLE FINGER : RING FINGER : (a) WRIST, (b) TOE, *(c) PINKY, (d) HAND

Novel Relevant

Furniture is eaten at the end of a meal.
APPETIZER : SOUP : SALAD : MAIN COURSE : *(a) TABLE, (b) BREAD, (c) MENU, (d) ENTREE
(Uncued condition: TABLE is replaced by DESSERT)

Novel Irrelevant

People always sleep standing up.
AWAKE : FATIGUED : DROWSY : ASLEEP : (a) ALERT, (b) TIRED, (c) CONFUSED, *(d) UNCONSCIOUS

Note: The correct answer is preceded by an asterisk in the table. Subjects were instructed to answer items as though the premise were true, assuming nothing else out of the ordinary other than the information in the premise.

Table 7

Reliabilities of Response Times

	Overall	Uncued	Cued
Analogies			
Item	.89	.89	.89
Subject	.89	.89	.89
Classifications			
Item	.87	.87	.88
Subject	.91	.91	.91
Series			
Item	.88	.85	.90
Subject	.92	.92	.92

Note: Only correct response times were used.

Table 8

Basic Statistics

Raw Scores

		Analogies		Classifications		Series		Overall	
		RT	ER	RT	ER	RT	ER	RT	ER
Familiar	Cued	3.91	2.4	3.81	2.6	5.15	5.4	4.29	3.5
Relevant	Uncued	3.61	0.7	4.48	6.1	5.55	11.5	4.55	6.1
Familiar	Cued	4.18	3.5	4.26	4.1	6.22	5.7	4.89	4.4
Irrelevant	Uncued	3.52	2.2	3.87	2.2	5.17	4.3	4.19	2.9
Novel	Cued	4.97	9.1	6.64	23.5	5.22	5.6	5.61	12.7
Relevant	Uncued	3.82	3.7	4.26	6.9	5.01	4.8	4.36	5.1
Novel	Cued	4.32	6.9	4.42	8.1	6.65	13.0	5.13	9.3
Irrelevant	Uncued	3.23	2.4	4.09	3.0	5.02	8.1	4.11	4.5

Note: Response times are expressed in seconds, error rates in percentages. Response times are for correct responses only.

Table 9

Analysis of Variance

Reaction Time (Correct Only)

Factors	df	MS	F
Test Type	2	302.1	106.42***
Test Type x Subject	118	2.8	
Cue	1	163.8	54.97***
Cue x Subject	59	3.0	
Novelty	1	36.7	53.84***
Novelty x Subject	59	0.7	
Relevance	1	5.0	10.88**
Relevance x Subject	59	0.5	
Test x Cue	2	1.2	1.15
Test x Cue x Subject	118	1.1	
Test x Novelty	2	18.5	28.29***
Test x Novelty x Subject	118	0.7	
Test x Relevance	2	39.7	51.10***
Test x Relevance x Subject	118	0.8	
Cue x Novelty	1	78.7	153.06***
Cue x Novelty x Subject	59	0.5	
Cue x Relevance	1	12.4	14.71***
Cue x Relevance x Subject	59	0.8	
Novelty x Relevance	1	18.3	25.50***
Novelty x Relevance x Subject	59	0.7	

Table 9 (continued)

Factors	df	MS	F
Test x Novelty x Relevance	2	16.3	24.91***
Test x Novelty x Relevance x Subject	118	0.7	
Cue x Novelty x Relevance	1	28.9	42.44***
Cue x Novelty x Relevance x Subject	59	0.7	
Test x Cue x Relevance x Novelty	2	20.8	23.95***
Test x Cue x Relevance x Novelty x Subject	118	0.9	

Table 9 (continued)

Analysis of Variance

Error Rate

Factor	df	MS	F
Test Type	2	.1760	23.75***
Test Type x Subject	118	.0074	
Cue	1	.2871	28.08***
Cue x Subject	59	.0102	
Novelty	1	.4899	89.59***
Novelty x Subject	59	.0055	
Relevance	1	.0879	16.69***
Relevance x Subject	59	.0053	
Test x Cue	2	.1416	10.40***
Test x Cue x Subject	118	.0068	
Test x Novelty	2	.0907	19.95***
Test x Novelty x Subject	118	.0045	
Test x Relevance	2	.1385	18.09***
Test x Relevance x Subject	118	.0076	
Cue x Novelty	1	.4109	44.01***
Cue x Novelty x Subject	59	.0093	
Cue x Relevance	1	.0046	.88
Cue x Relevance x Subject	59	.0052	
Novelty x Relevance	1	.0068	1.15
Novelty x Relevance x Subject	59	.0062	

Table 9 (continued)

Factor	df	MS	F
Test x Cue x Novelty	2	.0586	8.62***
Test x Cue x Novelty x Subject	118	.0068	
Test x Cue x Relevance	2	.0635	6.70**
Test x Cue x Relevance x Subject	118	.0095	
Test x Novelty x Relevance	2	.2327	29.78***
Test x Novelty x Relevance x Subject	118	.0078	
Cue x Novelty x Relevance	1	.1121	17.29***
Cue x Novelty x Relevance x Subject	59	.0064	

* $p < .05$

** $p < .01$

*** $p < .001$

Table 10

Difference Scores, Precued - Uncued,

for Reaction Time of Correct Responses (seconds)

and for Error Rate (percent)

	Analogies		Classifications		Series	
	RT	ER	RT	ER	RT	ER
Familiar Relevant	.30	1.7	-.67	-3.5	-.40	-6.1
Familiar Irrelevant	.66	1.3	.39	1.9	1.05	1.4
Novel Relevant	1.15	5.4	2.38	16.6	.21	0.8
Novel Irrelevant	1.09	4.5	.33	5.1	1.63	4.9

Table 11

Correlations of Response Times and Error Rates

With Ability Tests

	OVERALL		UNCUED		CUED	
	RT	ER	RT	ER	RT	ER
DAT Verbal Reasoning	-.21	-.58***	-.19	-.45***	-.23	-.57***
Cattell Abstract Reasoning	-.38**	-.06	-.32*	-.02	-.41**	-.08
Insight Problems	-.24	-.41***	-.20	-.36**	-.26*	-.39**
Crossing Out A's	-.09	-.06	-.16	-.10	-.04	.04
Extended Vocabulary (Level 3)	-.07	-.29*	-.10	-.26*	-.04	-.26*
Reasoning Factor	-.36**	-.46***	-.31*	-.36**	-.39**	-.45***
Verbal/Perceptual Factor	-.10	-.14	-.17	-.10	-.05	-.15

Note: Only correct response times were included.

* $p < .05$ ** $p < .01$ *** $p < .001$

Table 12

Correlations among Expertise-Related Variables

	Age	Exprts	Estlvl	MPtot	MPlast	Yrsplyd
Age		0.15	0.17	0.06	0.01	0.32
Exprts			0.77***	0.74***	0.76***	0.75***
Estlvl				0.53**	0.56***	0.75***
MPtot					0.95***	0.44*
MPlast						0.48**
Yrsplyd						

Abbreviations: Exprts: Rating on expertise scale

Estlvl: Self-estimated level of expertise

MPtot: Total number of master points accumulated

MPlast: Number of master points accumulated in past 2 years

Yrsplyd: Number of years bridge played prior to testing

* $p < .05$

** $p < .01$

*** $p < .001$

Table 13

Comparison of Expertise Groups

	Nonexperts		Experts	
Variable	Mean	Std	Mean	Std
Age	43.06	15.85	47.00	8.19
Exprts	2.06	0.57	6.51	1.85
Estlvl	4.09	1.75	7.21	0.92
MPtot	5.02	7.93	716.94	963.01
MPlast	2.48	3.27	186.53	203.07
Yrsplyd	2.54	1.81	21.88	8.67

Table 14

Spearman-Brown Corrected Split-Half Reliabilities of Dependent Variables

	Play Time	Tricks Won	Games Won
Overall	0.92	0.73	0.63
Normal games	0.85	0.55	0.51
Modified games	0.89	0.41	0.40

Table 15

Correlations Among Dependent Variables

	Play Time	Tricks Won	Games Won
Play Time			
Total		-0.11*	-0.07
Unchanged games		-0.13	-0.12
Modified games		-0.08	-0.01
Tricks Won			
Total			0.36***
Unchanged games			0.35***
Modified games			0.36***

* $p < .05$

*** $p < .001$

Table 16

Experiment 1

Basic Statistics for Nonexperts

Modification	Block	Play Time		Tricks Won		Games Won	
		Mean	Std	Mean	Std	Mean	Std
Name change	1	12.50	3.46	7.21	3.04	0.57	0.51
	2	14.85	4.52	6.20	2.73	0.67	0.49
	3	12.20	3.73	6.60	2.35	0.40	0.51
	4	10.06	2.93	7.00	3.21	0.60	0.51
Rank-order change	1	12.52	4.33	7.00	2.99	0.61	0.50
	2	13.47	4.53	6.17	3.29	0.56	0.51
	3	12.20	3.21	6.72	2.68	0.39	0.50
	4	10.03	3.17	6.94	2.75	0.44	0.51
Lead change	1	10.20	2.75	6.73	2.92	0.33	0.49
	2	10.66	2.47	6.47	2.20	0.47	0.52
	3	9.35	2.77	7.07	2.82	0.47	0.52
	4	7.36	2.04	6.73	3.60	0.60	0.51

Note: Means and Standard deviations of play times (per play) are expressed in seconds.

Table 17

Experiment 1

Basic Statistics for Experts

Modification	Block	Play Time		Tricks Won		Games Won	
		Mean	Std	Mean	Std	Mean	Std
Name change	1	10.04	2.92	6.22	3.34	0.72	0.46
	2	12.61	3.62	7.22	3.17	0.61	0.50
	3	9.69	2.65	7.00	2.93	0.50	0.52
	4	6.96	1.69	7.11	3.16	0.44	0.51
Rank-order change	1	8.64	2.48	6.24	2.79	0.62	0.50
	2	9.62	2.49	6.24	3.46	0.43	0.51
	3	8.71	2.59	7.24	3.03	0.52	0.51
	4	6.45	2.16	7.91	2.84	0.62	0.50
Lead change	1	9.42	2.99	6.13	3.44	0.60	0.51
	2	13.86	5.20	5.47	2.64	0.47	0.52
	3	10.89	3.94	7.53	2.70	0.40	0.51
	4	7.82	2.94	7.13	3.20	0.47	0.52

Note: Means and Standard deviations of play times (per play) are expressed in seconds.

Table 18

Basic Statistics for Unchanged Bridge Rules (Block 1)

	Nonexperts		Experts	
Variable	Mean	Std	Mean	Std
Play Time	11.86	3.01	9.33	2.35
Tricks Won	6.99	1.43	6.20	1.74
Games Won (%)	0.51	0.17	0.65	0.27

Note: Means and Standard deviations of play times (per response) are expressed in seconds.

Table 19

Change in Dependent Variables due to Game Modifications

Percentages

Expertise	Display Change	Play Time	Tricks Won	Games Won
Nonexperts	Name change	5.46	-5.94	8.33
	Rank-order change	4.94	-3.16	-20.83
	Lead change	-0.89	1.69	40.00
Experts	Name change	12.31	22.70	-1.39
	Rank-order change	6.53	30.95	-5.95
	Lead change	33.68	10.63	-16.67

Note: Means of changes in play time, number of tricks won, and number of games won are expressed in percent.

Table 20

Basic Statistics for Block 3

	Nonexperts		Experts	
Variable	Mean	Std	Mean	Std
Play Time	12.20	2.61	9.17	2.37
Tricks Won	6.67	1.29	7.13	1.73
Games Won (%)	0.39	0.29	0.51	0.29

Note: Means and Standard deviations of play times (per response) are expressed in seconds.

Table 21

Bid time (opening bids)

Alpha Reliabilities

	Bid Time	
	Item	Subject
Overall	0.93	0.93
Normal rules	0.88	0.93
Modified rules	0.91	0.90

Table 22

Bid Time (opening bids)

Raw Data

Modification	Block	Nonexperts		Experts	
		Mean	Std	Mean	Std
Name change	1	8.69	4.57	7.06	2.78
	2	17.25	11.22	18.61	12.74
	3	15.47	11.32	13.79	9.40
	4	9.78	5.73	8.21	4.29
Rank-order change	1	9.52	4.41	7.06	3.96
	2	14.86	9.43	14.32	11.20
	3	15.26	11.11	10.14	7.87
	4	10.07	5.84	8.48	7.75
Lead change	1	9.24	6.38	8.16	7.74
	2	12.75	9.14	13.78	11.57
	3	10.39	6.91	9.22	8.14
	4	9.92	6.14	7.66	5.97

Note: Means and Standard deviations of bid times (per bid) are expressed in seconds.

Table 23

Encoding

Alpha Reliabilities

	RT	Total	Correct
Overall			
Item	0.93	0.98	0.94
Subject	0.83	0.46	0.69
Normal Display			
Item	0.91	0.95	0.88
Subject	0.89	0.41	0.54
Modified Display			
Item	0.86	0.96	0.91
Subject	0.62	0.50	0.76

Abbreviations: RT: Reaction time

Total: Total number of cards written down

Correct: Number of correctly identified cards

Table 24

Correlations Among Dependent Variables

	RT	Total	Correct
RT			
Total	-	-0.13^{***}	-0.34^{***}
Normal Display	-	-0.10^{*}	-0.36^{***}
Modified Display	-	-0.17^{***}	-0.33^{***}
Total			
Total		-	0.74^{***}
Normal Display		-	0.70^{***}
Modified Display		-	0.78^{***}

Abbreviations: RT: Reaction time

Total: Total number of cards written down

Correct: Number of correctly identified cards

* $p < .05$

*** $p < .001$

Table 25

Experiment 3

Basic Statistics for Nonexperts

Modification	Block	RT		Total		Correct	
		Mean	Std	Mean	Std	Mean	Std
Name change	1	2.20	0.85	5.33	1.38	4.88	1.42
	2	1.89	0.75	4.92	1.34	4.67	1.49
	3	1.80	0.63	5.30	1.23	5.02	1.31
	4	1.72	0.53	5.12	1.22	4.92	1.29
Rank-order change	1	2.25	0.90	4.44	1.43	4.14	1.52
	2	2.21	0.75	4.29	1.34	4.25	1.42
	3	2.00	0.68	4.51	1.77	4.30	1.79
	4	1.78	0.58	4.36	1.50	4.29	1.47
Control	1	2.35	0.80	5.51	2.19	4.72	1.68
	2	1.86	0.63	5.70	1.86	5.65	1.79
	3	1.89	0.63	5.85	1.78	5.47	1.77
	4	1.87	0.74	5.35	1.37	4.96	1.58

Note: Means and Standard deviations of RT are expressed in seconds.

Table 26

Experiment 3

Basic Statistics for Experts

Modification	Block	RT		Total		Correct	
		Mean	Std	Mean	Std	Mean	Std
Name change	1	2.15	0.78	5.75	2.34	5.02	1.36
	2	1.89	0.81	5.22	1.76	5.08	1.70
	3	1.62	0.75	5.43	1.64	5.37	1.62
	4	1.45	0.51	5.95	1.38	5.54	1.64
Rank-order change	1	2.03	0.56	6.21	2.30	5.47	1.52
	2	2.12	0.53	5.52	1.95	5.07	1.87
	3	1.82	0.39	5.43	1.52	5.16	1.50
	4	1.68	0.48	5.29	1.93	5.00	1.56
Control	1	2.27	0.85	7.56	3.40	5.74	1.71
	2	1.94	0.57	7.35	3.28	5.78	1.80
	3	1.78	0.57	7.32	3.20	5.91	1.44
	4	1.72	0.59	7.00	3.03	5.38	1.43

Note: Means and Standard deviations of RT are expressed in seconds.

Table 27

Basic Statistics for Normal Display (Blocks 1 & 4)

	Nonexperts		Experts	
Variable	Mean	Std	Mean	Std
RT	2.02	0.41	1.87	0.39
Total	5.00	0.83	6.23	1.90
Correct	4.60	0.77	5.31	0.66

Note: Means and Standard deviations of RT are expressed in seconds.

Abbreviations:
- RT: Reaction time
- Total: Total number of cards written down
- Correct: Number of correctly identified cards

Table 28

Change in Dependent Variables due to Display Changes

Percentages

Expertise	Display Change	RT	Total	Correct
Novices	Name change	-3.64	-2.93	-1.97
	Rank-order change	9.18	-1.95	-6.81
	Control	-9.77	6.91	14.41
Experts	Name change	-3.40	-9.03	-10.43
	Rank-order change	7.86	-3.54	-2.09
	Control	-4.24	0.82	4.91

Note: Changes in RT, Total, and Correct are expressed in percent.

Abbreviations:
RT: Reaction time
Total: Total number of cards written down
Correct: Number of correctly identified cards

Table 29

Models for Nonentrenched Concepts

Model	Lexical familiarity: Familiar	Lexical familiarity: Unfamiliar	Conceptual naturalness
0	0	0	Natural
	0	0	Unnatural
1	-	-	Natural
	+	+	Unnatural
2	-	+	Natural
	-	+	Unnatural
3	--	0	Natural
	0	++	Unnatural
4	-	+	Natural
	+	-	Unnatural

Table 30

Definition of Concepts for Experiment 1

Season name	Beginning color	Ending color
Group 1: Leaves in New Haven--Natural concepts and familiar names		
Summer	Green	Green
Fall	Green	Brown
Winter	Brown	Brown
Spring	Brown	Green
Group 2: Leaves in Latzania--Natural concepts and unfamiliar names		
Soob	Green	Green
Trit	Green	Brown
Blen	Brown	Brown
Mave	Brown	Green
Group 3: Rocks on Kyron--Unnatural concepts and familiar names		
Summer	Blue	Blue
Fall	Blue	Orange
Winter	Orange	Orange
Spring	Orange	Blue
Group 4: Rocks on Kyron--Unnatural concepts and unfamiliar names		
Soob	Blue	Blue
Trit	Blue	Orange
Blen	Orange	Orange
Mave	Orange	Blue

Note: For sets 2, 3, and 4, the set of descriptions listed is only one of four possible counterbalanced versions that were used.

Table 31

Basic Statistics for Experiment 1

Group	Mean	Standard deviation over subjects	Standard deviation over item types
Response latencies			
1	3.83	.92	1.04
2	4.18	1.33	1.25
3	4.80	1.56	1.45
4	3.76	1.02	1.00
Error rates			
1	.06	.04	.05
2	.08	.04	.06
3	.09	.06	.09
4	.05	.03	.04

Note: Mean latencies are expressed in seconds. Each set of data is based upon 24 subjects and 36 item types.

Table 32

Indices of Model Fit for Experiment 1

		Group			
Index	Definition	1	2	3	4
R^2	Proportion of variance accounted for in latency data by the model	.93	.93	.92	.91
r(xx)	Item reliability	.95	.97	.97	.97
F reg	Regression F for R^2	32.7*	31.8*	28.8*	23.5*
df	Degrees of freedom for reg	5.12	5.12	5.12	5.12
r(Res)	Correlation between residuals of observed from predicted values for splits on random halves of subjects, corrected by Spearman-Brown formula	.80*	.90*	.80*	.74*
RMSD	Root-mean-square deviation of observed from predicted latencies	.31	.37	.45	.34

Note: Model fits are based on 18 data points.

Table 33

Latency Parameter Estimates for Experiment 1

Parameter name	Parameter description	Group 1	Group 2	Group 3	Group 4
AII	Assess initial state (, , 1)	1.70	1.90	1.18	1.38
CRS	Change in representation system (0-1)	0.83	0.88	1.22	0.72
CPS	Change in physical state (0-1)	1.30	1.64	2.15	1.44
UCVS	Unexpected lack of change (0-1)	0.85	0.85	1.36	0.74
JUS	Justification (0-1)	1.34	1.82	2.27	1.37

Note: All parameter estimates were significantly different from zero ($p < .05$).

Table 34

Correlations Between Projection Task Scores and Measured Abilities for Experiment 1

	Group			
	1	2	3	4
Induction				
Overall response latency	-.63**	-.14	-.57**	-.40**
Overall error rate	-.22	-.02	-.33	-.10
AII	-.23	-.11	-.26	-.31
CRS	-.46**	.19	-.49**	-.26
CPS	-.53**	-.24	-.56**	-.34
UCVS	-.29	.01	-.56**	-.40*
JUS	-.11	-.12	-.41*	-.24
Deduction				
Overall response latency	-.50**	.06	-.33	-.35
Overall error rate	-.14	-.03	-.31	-.62**
AII	-.28	-.20	-.12	-.40*
CRS	-.39*	.07	-.15	-.28
CPS	-.30*	-.24	.-49**	-.21
UCVS	-.17	.19	-.48**	-.26
JUS	.18	-.02	-.27	-.37*
Insight				
Overall response latency	-.22	-.27	-.53**	-.25

Note: All tests of significance are one-tailed.
* p < .05
** p < .01

Table 35

Definition of Concepts for Experiment 2

Group	Period name	Beginning shape or position	Ending shape or position
1: Sun in New Haven: Natural concepts and familiar names	Dawn	Below	Above
	Daytime	Above	Above
	Dusk	Above	Below
	Nighttime	Below	Below
2: Sun in Canada: Natural concepts and unfamiliar names	Trofar	Below	Above
	Kovit	Above	Above
	Bren	Above	Below
	Stobe	Below	Below
3: Mineral on Kyron: Unnatural concepts and familiar names	Dawn	Rectangular	Oval
	Daytime	Oval	Oval
	Dusk	Oval	Rectangular
	Nighttime	Rectangular	Rectangular
4: Mineral on Kyron: Unnatural concepts and unfamiliar names	Trofar	Rectangular	Oval
	Kovit	Oval	Oval
	Bren	Oval	Rectangular
	Stobe	Rectangular	Rectangular

Note: For sets 2, 3, and 4, the set of descriptions listed is only one of four possible counterbalanced versions that were used.

Table 36

Basic Statistics for Experiment 2

Group	Mean	Standard deviation over subjects	Standard deviation over item types
Response latencies			
1	5.42	1.44	1.45
2	5.54	1.91	1.37
3	6.92	2.56	1.67
4	5.36	1.36	1.36
Error rates			
1	.09	.07	.09
2	.08	.07	.07
3	.09	.08	.10
4	.05	.06	.06

Note: Mean latencies are expressed in seconds. Each set of data is based upon 20 subjects and 36 item types.

Table 37

Indices of Model Fit for Experiment 2

Index	Definition	Group 1	2	3	4
R^2	Proportion of variance accounted for in latency data by the model	.95	.91	.86	.93
r(xx)	Item reliability	.96	.97	.97	.95
F reg	Regression F for R^2	41.7*	23.6*	15.2*	29.8*
df	Degrees of freedom for reg	5.12	5.12	5.12	5.12
r(Res)	Correlation between residuals of observed from predicted values for splits on random halves of subjects, corrected by Spearman-Brown formula	.35	.67*	.71*	.74*
RMSD	Root-mean-square deviation of observed from predicted latencies	.39	.46	.65	.44

Note: Model fits are based on 18 data points.

* $p < .01$

Table 38

Latency Parameter Estimates for Experiment 2

Parameter name	Parameter description	Group 1	Group 2	Group 3	Group 4
AII	Assess initial state (, , 1)	2.31	2.05	1.97	1.66
CRS	Change in representation system (0-1)	1.33	0.94	1.81	1.28
CPS	Change in physical state (0-1)	1.60	1.39	1.49	1.73
UCVS	Unexpected lack of change (0-1)	1.48	1.18	1.28	1.47
JUS	Justification (0-1)	1.60	1.95	1.44	2.03

Note: All parameter estimates were significantly different from zero ($p < .05$).

Table 39

Correlations Between Projection Task Scores and Measured Abilities for Experiment 2

	Group			
	1	2	3	4
Induction				
Overall response latency	-.26	-.52**	-.52**	-.31
Overall error rate	-.37	-.52**	-.35	-.12
AII	-.17	-.10	.01	-.42
CRS	-.05	-.06	-.31**	-.06
CPS	-.35	-.58**	-.53**	-.15
UCVS	-.19	-.36	-.07	.24
JUS	-.17	-.31	-.54**	-.28
Deduction				
Overall response latency	-.25	-.36	-.67**	-.36
Overall error rate	-.26	-.61**	-.04	.11
AII	-.19	.03	-.32	-.55**
CRS	-.11	-.20	-.72**	-.21
CPS	-.02	-.20	-.14	.02
UCVS	.13	.08	.18	.28
JUS	.12	-.29	-.37	-.15*
Insight				
Overall response latency	-.18	-.53**	-.32	-.13

Note: All tests of significance are one-tailed.
* $p < .05$
** $p < .01$

Figure Caption

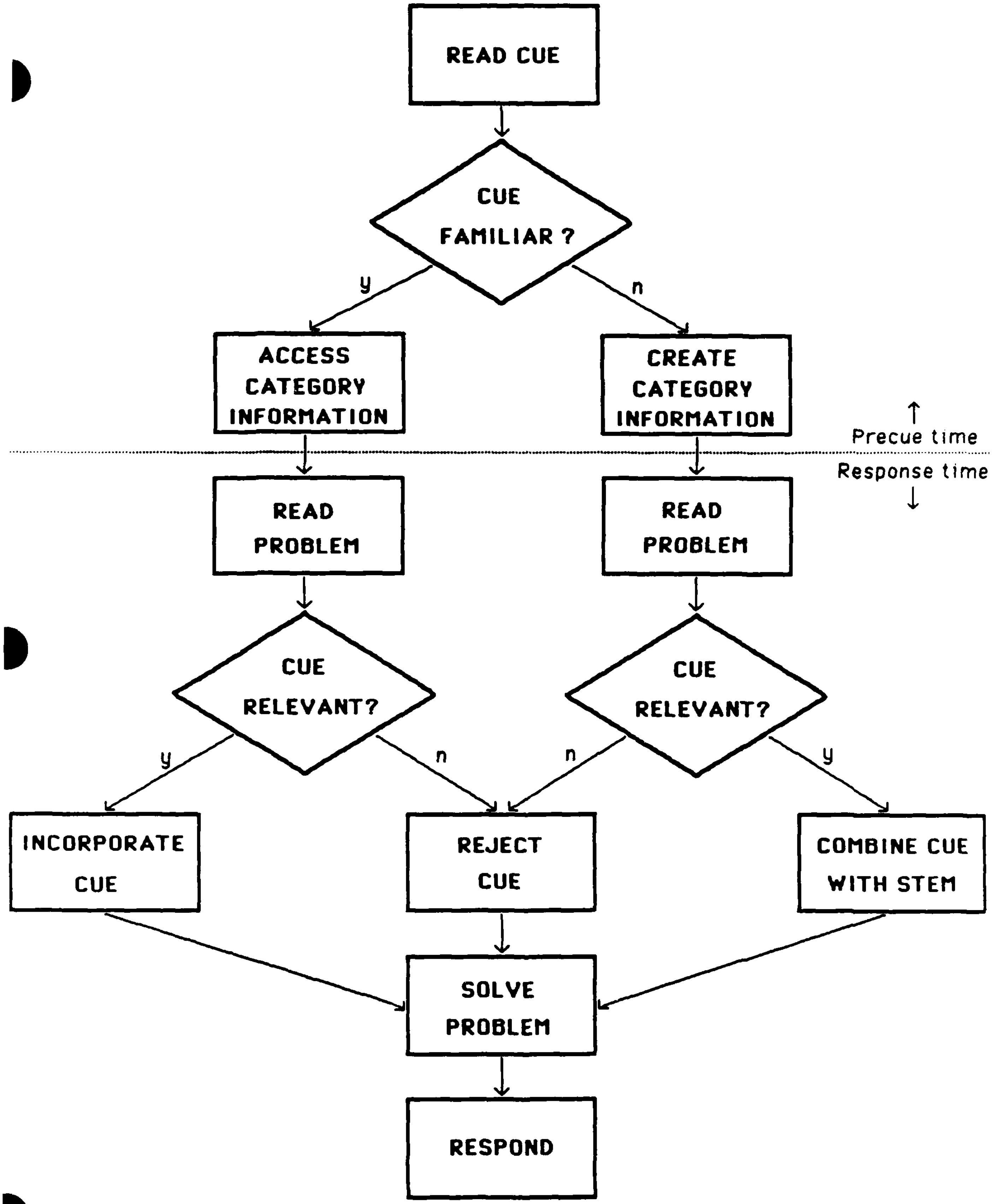

READ CUE
CUE FAMILIAR ?
y
n
ACCESS CATEGORY INFORMATION
CREATE CATEGORY INFORMATION
Precue time
Response time
READ PROBLEM
READ PROBLEM
CUE RELEVANT?
CUE RELEVANT?
y
n
n
y
INCORPORATE CUE
REJECT CUE
COMBINE CUE WITH STEM
SOLVE PROBLEM
RESPOND

Figure 1

GAME NO : 1

North

Declarer: E/W N/S
0 0

Contract: 2 ♡

West

5 ◇
4 ◇

East
♠ Q T 2
♡ K T 4 3
◇ T 8 6
♣ 7 3

South
♠ J 3
♡ A 8 7 5
◇ K J 9 3 2
♣ A 9

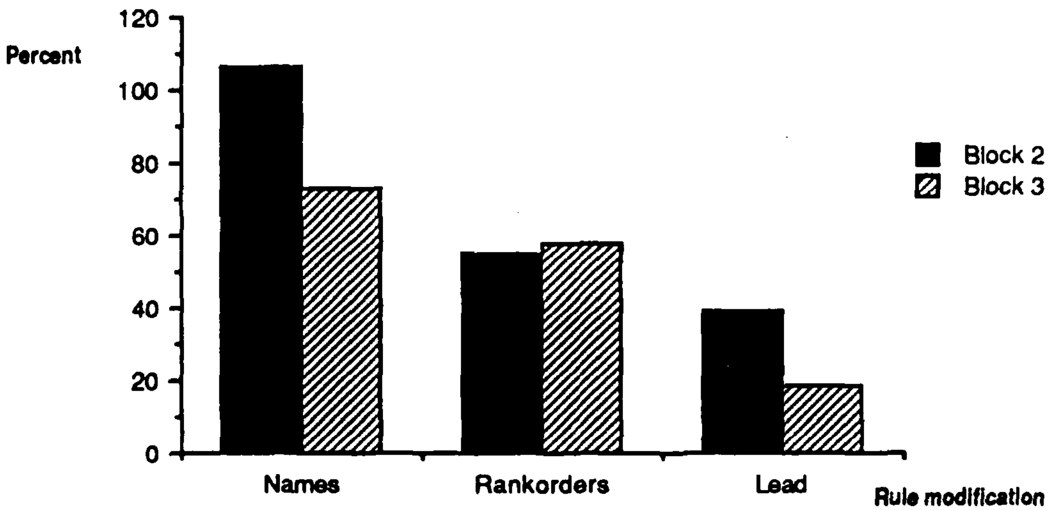
North	East	South	West
5 ◇	4 ◇		

South: ?

Card ranks: A,K,Q,J,T,9...2

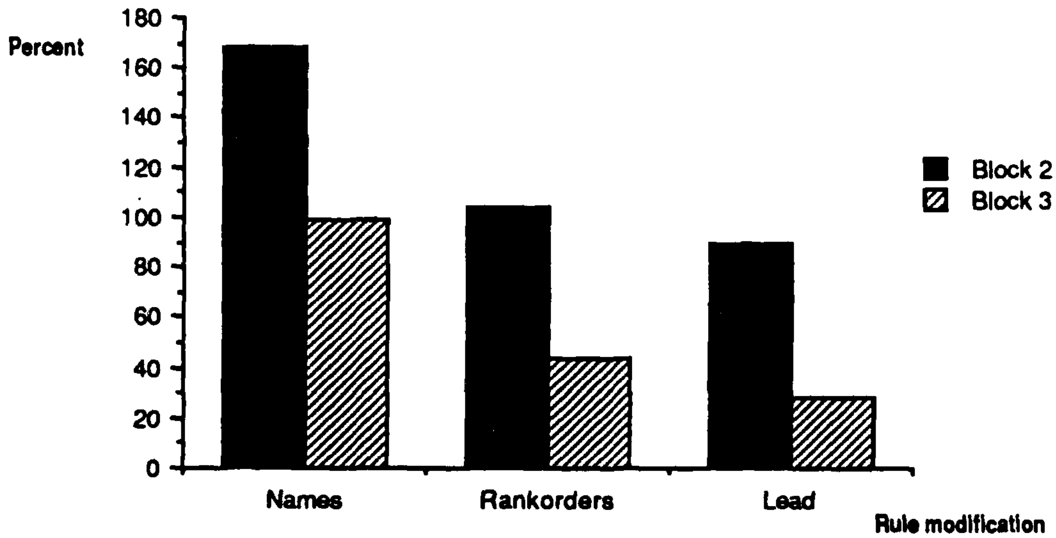
Suit ranks: Sp, He, Di,Cl

Figure 2



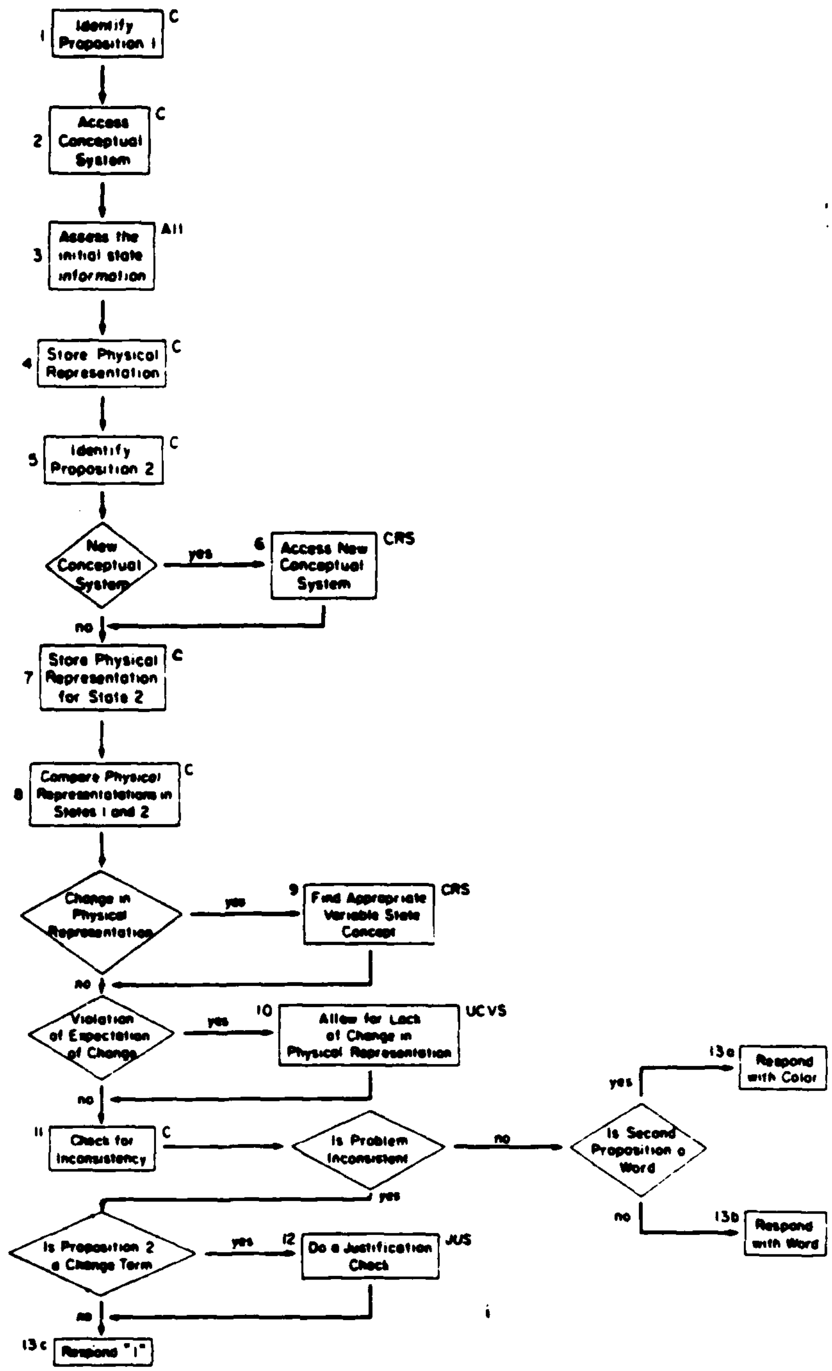
Percent
120
100
80
60
40
20
0
Block 2
Block 3
Names
Rankorders
Lead
Rule modification

Figure 3



Percent
180
160
140
120
100
80
60
40
20
0
Block 2
Block 3
Names
Rankorders
Lead
Rule modification

Figure 4



1 Identify Proposition 1 C
2 Access Conceptual System C
3 Assess the initial state information All
4 Store Physical Representation C
5 Identify Proposition 2 C
New Conceptual System
yes
6 Access New Conceptual System CRS
no
7 Store Physical Representation for State 2 C
8 Compare Physical Representatations in States 1 and 2 C
Change in Physical Representation
yes
9 Find Appropriate Variable State Concept CRS
no
Violation of Expectation of Change
yes
10 Allow for Lack of Change in Physical Representation UCVS
no
11 Check for Inconsistency C
Is Problem Inconsistent
no
Is Second Proposition a Word
yes
13a Respond with Color
no
13b Respond with Word
yes
Is Proposition 2 a Change Term
yes
12 Do a Justification Check JUS
no
13c Respond "I"

Figure 5

Section 2

Publications

1985

Berg, C. A., & Sternberg, R. J. (1985). Response to novelty: Continuity versus discontinuity in the developmental course of intelligence. In H. Reese (Ed.), Advances in child development and behavior (Vol. 19, pp. 2-47). New York: Academic Press.

Berg, C. A., & Sternberg, R. J. (1985). A triarchic theory of intellectual development during adulthood. Developmental Review, 5, 334-370.

Davidson, J. E., & Sternberg, R. J. (1985). Competence and performance in intellectual development. In E. Neimark, R. de Lisi, & J. H. Newman (Eds.), Moderators of competence (pp. 43-76). Hillsdale, NJ: Erlbaum.

Downing, C. J., Sternberg, R. J., & Ross, B. (1985). Multicausal inference: Evaluation of evidence in causally complex situations. Journal of Experimental Psychology: General, 114, 239-263.

Sternberg, R. J. (1985). All's well that ends well, but it's a sad tale that begins at the end: A reply to Glaser. American Psychologist, 40, 571-573.

Sternberg, R. J. (1985). Applying componential analysis to the study of individual differences in cognitive skills. In C. Reynolds & V. Willson (Eds.), Methodological and statistical advances in the study of individual differences. New York: Plenum.

Sternberg, R. J. (1985). Beyond IQ: A triarchic theory of human intelligence. New York: Cambridge University Press.

Sternberg, R. J. (1985). The black-white differences and Spearman's g: Old wine in new bottles that still doesn't taste good. Behavioral and Brain Sciences, 8, 244.

Sternberg, R. J. (1985). Cognitive approaches to intelligence. In B. Wolman (Ed.), Handbook of intelligence. New York: Wiley.

Sternberg, R. J. (1985). Componential analysis: A recipe. In D. K. Detterman (Ed.), Current topics in human intelligence (Vol. 1, pp. 179-201). Norwood, NJ: Ablex.

Sternberg, R. J. (1985). Controlled versus automatic processing: A reply to Fodor. Behavioral and Brain Sciences, 8, 32-33.

Sternberg, R. J. (1985). Critical thinking: Its nature, measurement, and improvement. In F. R. Link (Ed.), Essays on the intellect (pp. 45-65). Alexandria, VA: Association for Supervision and Curriculum Development.

Sternberg, R. J. (1985). For the best on how to test: Review of Essentials of psychological testing (4th ed.), by Lee J. Cronbach. Contemporary Psychology, 30, 377-378.

Sternberg, R. J. (1985). Human intelligence: The model is the message. Science, 230, 1111-1118.

Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. Journal of Personality and Social Psychology, 49, 607-627.

Sternberg, R. J. (1985). Instrumental and componential approaches to the training of intelligence. In S. Chipman, J. Segal, & R. Glaser (Eds.), Thinking and learning skills: Current research and open questions (Vol. 2, pp. 215-243). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1985). Pretty close to ideal, anyway: A review of The ideal problem solver. Human Intelligence Newsletter, 6, Spring, 10.

Sternberg, R. J. (1985). Review of Meichenbaum, Burland, Grusen, and Cameron's "Metacognitive assessment." In S. Yussen (Ed.), The growth of reflection in the child (pp. 31-35). New York: Academic Press.

Sternberg, R. J. (1985). Review of Nonverbal Test of Cognitive Skills. In J. V. Mitchell (Ed.), The ninth mental measurements yearbook. Lincoln, NE: University of Nebraska Press.

Sternberg, R. J. (1985). Review of Test of Cognitive Skills. In J. V. Mitchell (Ed.), The ninth mental measurements yearbook. Lincoln, NE: University of Nebraska Press.

Sternberg, R. J. (1985). Tacit agreements between authors and editors. Behavioral and Brain Sciences, 8, 746-747.

Sternberg, R. J. (1985). Teaching critical thinking, Part 1: Are we making critical mistakes? Phi Delta Kappan, 67, 194-198.

Sternberg, R. J. (1985). Teaching critical thinking, Part 2: Possible solutions. Phi Delta Kappan, 67, 277-280.

Sternberg, R. J., & Baron, J. B. (1985). A statewide approach to measuring critical thinking skills. Educational Leadership, 43, 40-43.

Sternberg, R. J., & Caruso, D. (1985). Practical modes of knowing. In E. Eisner (Ed.), Learning the ways of knowing (pp. 133-158). Chicago, IL: University of Chicago Press.

Sternberg, R. J., & Davidson, J. E. (1985). Cognitive development in the gifted and talented. In F. D. Horowitz & M. O'Brien (Eds.), The gifted and talented: A developmental perspective. Washington, DC: American Psychological Association.

Sternberg, R. J., & McNamara, T. P. (1985). The representation and processing of information in real-time verbal comprehension. In S. E. Embretson (Ed.), Test design: Contributions from psychology, education, and psychometrics (pp. 21-43). New York: Academic Press.

Sternberg, R. J., & Smith, C. (1985). Social intelligence and decoding skills in nonverbal communication. Social Cognition, 2, 168-192.

Sternberg, R. J., & Spear, L. C. (1985). A triarchic theory of mental retardation. In N. Ellis & N. Bray (Eds.), International review of research in mental retardation (Vol. 13, pp. 301-326). New York: Academic Press.

Wagner, R. K., & Sternberg, R. J. (1985). Practical intelligence in real-world pursuits: The role of tacit knowledge. Journal of Personality and Social Psychology, 49, 436-458.

1986

Davidson, J. E., & Sternberg, R. J. (1986). What is insight? Educational Horizons, 64, 177-179.

Dillon, R. F., & Sternberg, R. J. (Eds.) (1986). Cognition and instruction. Orlando, FL: Academic Press.

Marr, D. B., & Sternberg, R. J. (1986). Analogical reasoning with novel concepts: Differential attention of intellectually gifted and nongifted children to relevant and irrelevant novel stimuli. Cognitive Development, 1, 53-72.

Spear, L. C., & Sternberg, R. J. (1986). Cognitive assessment with disabled readers. Special Services in the Schools, 1985-1986, 2 (2/3), 71-84.

Sternberg, R. J. (Ed.) (1986). Advances in the psychology of human intelligence (Vol. 3). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1986). Alternatives to the triarchic theory of intelligence. Behavioral and Brain Sciences, 9, 581-582.

Sternberg, R. J. (1986). Cognition and instruction: Why the marriage sometimes ends in divorce. In R. F. Dillon & R. J. Sternberg (Eds.), Cognition and instruction (pp. 375-382). Orlando, FL: Academic Press.

Sternberg, R. J. (1986). A framework for understanding conceptions of intelligence. In R. J. Sternberg & D. K. Detterman (Eds.), What is intelligence? Contemporary viewpoints on its nature and definition (pp. 3-15). Norwood, NJ: Ablex.

Sternberg, R. J. (1986). The future of intelligence testing. Educational Measurement: Issues and Practice, 5(3), 19-22.

Sternberg, R. J. (1986). GENECES: A framework for intellectual abilities and theories of them. Intelligence, 10, 239-250.

Sternberg, R. J. (1986). Haste makes waste versus a stitch in time? A reply to Vernon, Nador, and Kantor. Intelligence, 10, 265-270.

Sternberg, R. J. (1986). How to prepare for the Miller Analogies Test (4th ed.). Woodbury, NY: Barron's Educational Series, Inc.

Sternberg, R. J. (1986). Identifying the gifted through IQ: Why a little bit of knowledge is a dangerous thing. Roeper Review, 8, 143-147.

Sternberg, R. J. (1986). In defense of 'critical thinking' programs. Education Week, 6(6), 19.

Sternberg, R. J. (1986). Inside intelligence. American Scientist, 74, 137-143.

Sternberg, R. J. (1986). Instructor's manual to accompany Intelligence applied: Understanding and increasing your intellectual skills. San Diego, CA: Harcourt, Brace, Jovanovich.

Sternberg, R. J. (1986). Intelligence applied: Understanding and increasing your intellectual skills. San Diego, CA: Harcourt, Brace, Jovanovich.

Sternberg, R. J. (1986). Intelligence is mental self-government. In R. J. Sternberg & D. K. Detterman (Eds.), What is intelligence? Contemporary viewpoints on its nature and definition (pp. 141-148). Norwood, NJ: Ablex.

Sternberg, R. J. (1986). Intelligence, wisdom, and creativity: Three is better than one. Educational Psychologist, 21, 175-190.

Sternberg, R. J. (1986). Introduction: The nature and scope of practical intelligence. In R. J. Sternberg & R. K. Wagner (Eds.), Practical intelligence: Nature and origins of competence in the everyday world (pp. 1-10). New York: Cambridge University Press.

Sternberg, R. J. (1986). A perspective on the prospects for Human intelligence: Perspectives and prospects, by R. Kail & J. W. Pellegrino. Contemporary Psychology, 31, 18-19.

Sternberg, R. J. (1986). Time capsules are not a panacea: A reply to Vernon. Intelligence, 10, 277-279.

Sternberg, R. J. (1986). Toward a unified theory of human reasoning. Intelligence, 10, 281-315.

Sternberg, R. J. (1986). A triarchic theory of human intelligence. In S. E. Newstead, S. H. Irvine, & P. L. Dann (Eds.), Human assessment: cognition and motivation (pp. 43-47). Dordrecht, The Netherlands: Martinus Nijhoff.

Sternberg, R. J. (1986). A triarchic theory of intellectual giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), Conceptions of giftedness (pp. 223-243). New York: Cambridge University Press.

Sternberg, R. J. (1986). What is adaptive? Behavioral and Brain Sciences, 9, 207-208.

Sternberg, R. J. (1986). What would better intelligence tests look like? In Measures in the college admissions process: A College Board Colloquium (pp. 146-150). New York: College Board.

Sternberg, R. J., & Berg, C. A. (1986). Quantitative integration: Definitions of intelligence: A comparison of the 1921 and 1986 symposia. In R. J. Sternberg & D. K. Detterman (Eds.), What is intelligence? Contemporary viewpoints on its nature and definition (pp. 155-162). Norwood, NJ: Ablex.

Sternberg, R. J., & Bhana, K. (1986). Synthesis of research on the effectiveness of intellectual skills programs: Snake-oil remedies or miracle cures? Educational Leadership, 44(2), 60-67.

Sternberg, R. J., & Davidson, J. E. (Eds.) (1986). Conceptions of giftedness. New York: Cambridge University Press.

Sternberg, R. J., & Davidson, J. E. (1986). Conceptions of giftedness: A map of the terrain. In R. J. Sternberg & J. E. Davidson (Eds.), Conceptions of giftedness (pp. 3-18). New York: Cambridge University Press.

Sternberg, R. J., & Detterman D. K. (Eds.) (1986). What is intelligence? Contemporary viewpoints on its nature and definition. Norwood, NJ: Ablex.

Sternberg, R. J., & Suben, J. (1986). The socialization of intelligence. In M. Perlmutter (Ed.), Perspectives on intellectual development (Vol. 19): Minnesota symposia on child psychology (pp. 201-235). Hillsdale, NJ: Erlbaum.

Sternberg, R. J., & Wagner, R. K. (Eds.) (1986). Practical intelligence: Nature and origins of competence in the everyday world. New York: Cambridge University Press.

Tetewsky, S. J., & Sternberg, R. J. (1986). Conceptual and lexical determinants of nonentrenched thinking. Journal of Memory and Language, 25, 202-225.

Wagner, R. K., & Sternberg, R. J. (1986). Tacit knowledge and intelligence in the everyday world. In R. J. Sternberg & R. K. Wagner (Eds.), Practical intelligence: Nature and origins of competence in the everyday world (pp. 51-83). New York: Cambridge University Press.

1987

Baron, J. B., & Sternberg, R. J. (Eds.) (1987). Teaching thinking skills: Theory and practice. New York: Freeman.

Kaye, D. B., Sternberg, R. J., & Fonseca, L. (1987). Verbal comprehension: The lexical decomposition strategy to define unfamiliar words. Intelligence, 11, 1-20.

Kolligian, J., Jr., & Sternberg, R. J. (1987). Another look at intelligence and learning disabilities: A reply to Reynolds' 'rap'. Journal of Learning Disabilities, 20(6), 325-326.

Kolligian, J., Jr., & Sternberg, R. J. (1987). Intelligence, information processing, and specific learning disabilities: A triarchic synthesis. Journal of Learning Disabilities, 20, 8-17.

Marr, D. B., & Sternberg, R. J. (1987). The role of mental speed in intelligence: A triarchic perspective. In P. A. Vernon (Ed.), Speed of information processing and intelligence (pp. 271-294). Norwood, NJ: Ablex.

Sincoff, J., & Sternberg, R. J. (1987). Two faces of verbal ability. Intelligence, 11, 263-276.

Spear, L. C., & Sternberg, R. J. (1987). An information-processing framework for understanding reading disabilities. In S Ceci (Ed.), Handbook of cognitive, social, and neuropsychological aspects of learning disabilities (Vol. 2, pp. 3-31). Hillsdale, NJ: Erlbaum.

Spear, L. C., & Sternberg, R. J. (1987). Teaching styles: Staff development for teaching thinking. Journal of Staff Development, 8(3), 35-39.

Sternberg, R. J. (1987). Behavior genetics moves beyond percentages--At last. Behavioral and Brain Sciences, 10, 40.

Sternberg, R. J. (1987). Coping with novelty and human intelligence. In P. Morris (Ed.), Modelling cognition (pp. 57-91). London: Wiley.

Sternberg, R. J. (1987). A day at developmental downs: Sportscast #2-Neo-Piagetian theories of cognitive development. International Journal of Psychology, 22(5/6), 507-529.

Sternberg, R. J. (1987). Difficulties in comparing intelligence across species. Behavioral and Brain Sciences, 10(4), 679-680.

Sternberg, R. J. (1987). Down with paper clips! A review of V. John-Steiner, Notebooks of the mind: Explorations of thinking. Contemporary Psychology, 32, 339-340.

Sternberg, R. J. (1987). Five ways to think about thinking skills. Instructor, Special Issue, Fall, 32-33.

Sternberg, R. J. (1987). Gee, there's more than g: A critique of Arthur Jensen's views on intelligence. In S. Modgil & C. Modgil (Eds.), Arthur Jensen: Consensus and controversy (pp. 237-249). Barcombe, England: Falmer Press.

Sternberg, R. J. (1987). Implicit theories: An alternative to modeling cognition and its development. In J. Bisanz, C. Brainerd, & R. Kail (Eds.), Formal methods in developmental psychology: Progress in cognitive development research (pp. 155-192). New York: Springer-Verlag.

Sternberg, R. J. (1987). Intelligence. In R. L. Gregory (Ed.), The Oxford companion to the mind (pp. 375-379). Oxford, England: Oxford University Press.

Sternberg, R. J. (1987). Intelligence and cognitive style. In R. E. Snow & M. J. Farr (Eds.), Aptitude, learning, and instruction, Vol. 3: Conative and affective process analyses (pp. 77-97). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1987). Most vocabulary is learned from context. In M. G. McKweon & M. E. Curtis (Eds.), The nature of vocabulary acquisition (pp. 89-105). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1987). The psychology of verbal comprehension. In R. Glaser (Ed.), Advances in instructional psychology (Vol. 3, pp. 97-151). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1987). Questioning and intelligence. Questioning Exchange, 1, 11-14.

Sternberg, R. J. (1987). Questions and answers about the nature and teaching of thinking skills. In J. B. Baron & R. J. Sternberg (Eds.), Teaching thinking: Theory and practice (pp. 251-259). New York: Freeman.

Sternberg, R. J. (1987). Review of B. Presseisen, Thinking skills. Educational Horizons, 65, 151.

Sternberg, R. J. (1987). Second game: A school's-eye view of intelligence. In J. A. Langer (Ed.), Language, literacy, and culture: Issues of society and schooling (pp. 23-48). Norwood, NJ: Ablex.

Sternberg, R. J. (1987). Stalking a moving target. Review of R. Nickerson, D. Perkins, & E. Smith, The teaching of thinking. Contemporary Psychology, 32, 322-323.

Sternberg, R. J. (1987). Synopsis of a triarchic theory of human intelligence. In S. H. Irvine & S. E. Newstead (Eds.), Intelligence and cognition: Contemporary frames of reference (pp. 141-176). Dordrecht, The Netherlands: Martinus Nijhoff.

Sternberg, R. J. (1987). Teaching critical thinking: Eight easy ways to fail before you begin. Phi Delta Kappan, 68, 456-459.

Sternberg, R. J. (1987). Teaching intelligence: The application of cognitive psychology to the improvement of intellectual skills. In J. B. Baron & R. J. Sternberg (Eds.), Teaching thinking skills: Theory and practice (pp. 182-218). New York: Freeman.

Sternberg, R. J. (1987). Teaching intelligence: A triarchic model. In D. N. Perkins, J. Lochhead, & J. C. Bishop (Eds.), Thinking: The second international conference (pp. 53-60). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1987). Three portraits of intellectual giftedness. Gifted Students Institute Quarterly, 12, A1-A5.

Sternberg, R. J. (1987). The triarchic theory of human intelligence: A framework for the understanding, investigation, testing, and training of intelligence. In J. Richardson, M. W. Eysenck, & D. W. Piper (Eds.), Student learning: Research in education and cognitive psychology (pp. 357-374). London: SRHE & The Open University Press.

Sternberg, R. J. (1987). A unified theoretical perspective on autism. In D. Cohen & A. M. Donnellan (Eds.), Handbook of autism and pervasive developmental disorders (pp. 690-696). New York: Wiley.

Sternberg, R. J. (1987). A unified theory of intellectual exceptionality. In J. D. Day & J. G. Borkowski (Eds.), Intelligence and exceptionality: New directions for theory, assessment, and instructional practices (pp. 135-172). Norwood, NJ: Ablex.

Sternberg, R. J. (1987). The uses and misuses of intelligence testing. Education Week, Sept. 23, 28, 22.

Sternberg, R. J., & Berg, C. A. (1987). What are theories of adult intellectual development theories of? In C. Schooler & K. Warner Schaie (Eds.), Cognitive functioning and social structure over the life course (pp. 3-23). Norwood, NJ: Ablex.

Sternberg, R. J., & Davidson, J. E. (1987). Teaching thinking to college students: Some lessons learned from experience. Teaching Thinking and Problem Solving, 9, May/June, 1-2, 10-11.

Sternberg, R. J., & Dobson, D. M. (1987). Resolving interpersonal conflicts: An analysis of stylistic consistency. Journal of Personality and Social Psychology, 52, 794-812.

Wagner, R. K., & Sternberg, R. J. (1987). Executive control in reading comprehension. In B. K. Britton & S. M. Glynn (Eds.), Executive control processes in reading (pp. 1-21). Hillsdale, NJ: Erlbaum.

1988

Okagaki, L., & Sternberg, R. J. (1988). Unwrapping giftedness. In G. Kanselaar, J. L. van der Linden, & A. Pennings (Eds.), Individual differences in giftedness: Identification and education (pp. 30-45). Louvain, Belgium: Acco.

Sincoff, J. B., & Sternberg, R. J. (1988). The development of verbal fluency abilities and strategies in elementary-school-aged children. Developmental Psychology, 24, 646-653.

Sternberg, R. J. (Ed.) (1988). Advances in the psychology of human intelligence (Vol. 4). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1988). Beyond IQ testing. National Forum, Spring, 8-11.

Sternberg, R. J. (1988). The development of intellectual styles. Teaching Thinking and Problem Solving, 10(2), 1-4.

Sternberg, R. J. (1988). GENECES: A rationale for the construct validation of theories and tests of intelligence. In H. Wainer & H. I. Braun (Eds.), Test validity (pp. 61-75). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1988). The intellect: Three portraits unveiled. Gifted Children Monthly, 9, 1-3.

Sternberg, R. J. (1988). Intellectual development: Psychometric and information-processing approaches. In M. Bornstein & M. Lamb (Eds.), Developmental psychology: An advanced textbook (2nd ed., pp. 261-295). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1988). Intelligence. In R. J. Sternberg & E. E. Smith (Eds.), The psychology of human thought (pp. 267-308). New York: Cambridge University Press.

Sternberg, R. J. (1988). Intelligence is more than a matter of "How much?" Human Intelligence International Newsletter, 8/4, 2-4.

Sternberg, R. J. (1988). Lessons from the life span: What theorists of intellectual development among children can learn from their counterparts studying adults. In E. M. Hetherington, R. M. Lerner, & M. Perlmutter (Eds.), Child development in life-span perspective (pp. 259-275). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (1988). The male/female difference is there: Should we care? Behavioral and Brain Sciences, 11(2), 210-211.

Sternberg, R. J. (Ed.) (1988). The nature of creativity. New York: Cambridge University Press.

Sternberg, R. J. (1988). A practical approach to practical intelligence. Contemporary Psychology, 33, 504-505.

Sternberg, R. J. (1988). The psychologist's companion (2nd ed.). New York: Cambridge University Press.

Sternberg, R. J. (1988). Questioning cognitive psychology. Contemporary Psychology, 33, 206-207.

Sternberg, R. J. (1988). Review of The long-term retention of knowledge and skills: A cognitive and instructional perspective. Human Intelligence International Newsletter, 8/4, 10.

Sternberg, R. J. (1988). Review of H. Spitz, The raising of intelligence: A selected history of attempts to raise retarded intelligence. American Journal of Psychology, 101, 142-145.

Sternberg, R. J. (1988). Survival of the fittest in theories of creativity. Journal of Social and Biological Structures, 11, 154-155.

Sternberg, R. J. (1988). A three-facet model of creativity. In R. J. Sternberg (Ed.), The nature of creativity (pp. 125-147). New York: Cambridge University Press.

Sternberg, R. J. (1988). The triarchic mind: A new theory of human intelligence. New York: Viking.

Sternberg, R. J., & Martin, M. (1988). When teaching thinking does not work, what goes wrong? Teachers College Record, 89, 555-578.

Sternberg, R. J., & Smith, E. E. (Eds.) (1988). The psychology of human thought. New York: Cambridge University Press.

Tardif, T., & Sternberg, R. J. (1988). Conclusions: What do we know about creativity? In R. J. Sternberg (Ed.), The nature of creativity (pp. 429-440). New York: Cambridge University Press.

In press

Ackerman, P., Sternberg, R. J., & Glaser, R. (Eds.) (in press). Learning and individual differences. New York: Freeman.

Kalmar, D. A., & Sternberg, R. J. (in press). Theory knitting: An integrative approach to theory development. Philosophical Psychology.

Kolligian, J., Jr., & Sternberg, R. J. (Eds.) (in press). Perceptions of competence and incompetence. New Haven, CT: Yale University Press.

Frensch, P. A., & Sternberg, R. J. (in press). Expertise and intelligent thinking: When is it worse to know better? In R. J. Sternberg (Ed.), Advances in the psychology of human intelligence (Vol. 5). Hillsdale, NJ: Erlbaum.

Sincoff, J., & Sternberg, R. J. (in press). The development of cognitive skills. In A. M. Colley & J. R. Beech (Eds.), The acquisition and performance of cognitive skills. Chichester, England: Wiley.

Sternberg, R. J. (Ed.) (in press). Advances in the psychology of human intelligence (Vol. 5). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (in press). Applying cognitive theory to the testing and teaching of intelligence. Applied Cognitive Psychology.

Sternberg, R. J. (in press). Behind closed doors: Unlocking the mysteries of human intelligence. In J. Brockman (Ed.), The reality club II. New York: LYNX.

Sternberg, R. J. (in press). CAT: A program for comprehensive abilities testing. Washington, DC: Commission on Testing and Public Policy.

Sternberg, R. J. (in press). Cognitive theory and psychometrics. In R. K. Hambleton & J. Zaal (Eds.), Psychological testing. Amsterdam: North-Holland.

Sternberg, R. J. (in press). Explaining away intelligence: A reply to Howe. British Journal of Psychology.

Sternberg, R. J. (in press). Intelligence. Encyclopedia Britannica. Chicago, IL: Britannica Inc.

Sternberg, R. J. (in press). Intelligence, wisdom, and creativity: Their natures and interralationships. In R. Linn (Ed.), Intelligence: Measurement, theory, and public policy. Champaign, IL: University of Illinois Press.

Sternberg, R. J. (in press). Mental self-government: A theory of intellectual styles and their development. Human Development.

Sternberg, R. J. (in press). Metaphors of the mind: Conceptions of the nature of intelligence. New York: Cambridge University Press.

Sternberg, R. J. (in press). Preface: The who, what, where, when, why, and how of theory and research on cognitive instruction. In B. F. Jones & L. Idol (Eds.), Dimensions of thinking and cognitive instruction (Vol. 2). Hillsdale, NJ: Erlbaum.

Sternberg, R. J. (in press). Prototypes of competence and incompetence. In J. Kolligian & R. J. Sternberg (Eds.), Perceptions of competence and incompetence. New Haven, CT: Yale University Press.

Sternberg, R. J. (in press). Smart love. New York: Bantam.

Sternberg, R. J. (in press). Sternberg Multidimensional Abilities Test. San Antonio, TX: The Psychological Corporation.

Sternberg, R. J. (in press). A triarchic view of intelligence in cross-cultural perspective. In S. Irvine (Ed.), The cultural context of human abilities. New York: Cambridge University Press.

Sternberg, R. J. (in press). What we don't know that we do know: A review of Nonconscious social information processing, by Pawel Lewicki. Journal of Social and Biological Structures.

Sternberg, R. J. (Ed.) (in press). Wisdom: Its nature, origins, and development. New York: Cambridge University Press.

Sternberg, R. J., & Davidson, J. E. (in press). A four-prong model for intellectual skills development. Journal of Research and Development in Education.

Sternberg, R. J., & Frensch, P. A. (in press). A balance-level theory of intelligent thinking. Zeitschrift fur Padagogische Psychologie (German Journal of Educational Psychology).

Sternberg, R. J., & Frensch, P. A. (Eds.) (in press). Complex problem solving: Principles and mechanisms. Hillsdale, NJ: Erlbaum.

Sternberg, R. J., & Frensch, P. (in press). Intelligence and cognition. In M. Eysenck (Ed.), International review of cognitive psychology (Vol. 1). Chichester, England: Wiley.

Sternberg, R. J., & Gastel, J. (in press). If dancers ate their shoes: Inductive reasoning with factual and counterfactual premises. Memory and Cognition.

Sternberg, R. J., & Wagner, R. K. (in press). The fate of the trait: A reply to Cantor and Kihlstrom. In R. S. Wyer, & T. K Srull (Eds.), Advances in social cognition (Vol. 2). Hillsdale, NJ: Erlbaum.

Sternberg, R. J., & Wagner, R. K. (in press). Individual differences in practical knowledge and its acquisition. In P. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), Learning and individual differences. New York: Freeman.

Wagner, R. K., & Sternberg, R. J. (in press). Tacit knowledge in managerial success. Journal of Business and Psychology.

Wagner, R. K., & Sternberg, R. J. (in press). Tacit knowledge: Its uses in identifying, assessing, and developing managerial talent. In J. Jones, B. Steffy, & D. Bray (Eds.), Applying psychology in business: The manager's handbook. New York: Human Sciences Press.

Williams, W. M., & Sternberg, R. J. (in press). Group intelligence: Why some groups are better than others. Intelligence.

Section 3

Presentations

October, 1985
- An informal presentation at the Psychological Corporation
- A colloquium at the University of Delaware
- An invited address to the Association for Gifted and Talented Education at the Concord Hotel in New York State
- A colloquium at Harvard University
- A colloquium at Lehigh University
- The keynote address to the New York State Conference on Thinking in Albany
- A workshop in Vernon, Connecticut
- A talk at a conference at Stanford University
- The keynote address at the Michigan State Conference on Thinking in Detroit.
- A colloquium at the University of Michigan

November, 1985
- A talk to a group of administrators at Southern Connecticut State University
- The keynote address at the National Association for Gifted Children
- An address at a week-long conference on cognition in Madrid, Spain
- A talk at the meeting of the Psychonomic Society in Boston
- The keynote address at the meeting of the Pennsylvania Association for Supervision and Curriculum Development

December, 1985
- An invited talk to the Washington State Conference on Thinking in Seattle
- A colloquium at the Center for Creative Leadership in Greensboro
- An invited talk to teachers in Glastonbury, Connecticut
- An invited talk to teachers in Southington, Connecticut
- Two colloquia at Vanderbilt University

January, 1986
- A workshop on thinking skills to the West Hartford School System
- A colloquium on intelligence to the Bank Street College of Education
- A Keynote address on teaching thinking to the Pennsylvania Association for Supervision and Curriculum Development

February, 1986
- A colloquium on intelligence at the University of Texas at Austin
- A colloquium on intelligence at Rice University in Houston
- A talk on practical intelligence for business writers in New York
- A talk on practical intelligence at Montclair State College in Upper Montclair, New Jersey
- Talks at Lake Washington School District, Washington, and at the annual meeting of the Association for Supervision and Curriculum Development

March, 1986

A talk at Sacred Heart University, Fairfield, CT
A televised talk to teachers in the State of Virginia
A workshop in Plainville, CT
An invited address on intelligence as mental self-government at Gatlingberg conference

April, 1986

Keynote addresses and workshops at Cheshire (CT) School District and Richmond School District (BC, Canada)
Keynote addresses at three conferences on the gifted at Michigan State University, Rutgers (New Jersey State Conference on the Gifted), and BOCES of New York
A colloquium at Columbia University Teachers College
A colloquium at Western Connecticut State University

May, 1986

A colloquium at Loyola University (Chicago)
A talk at the New School for Social Research
Talks at thinking skills conferences sponsored by the City of New York, South Dakota State University, and the International Foundation for Learning (Vancouver, BC)

June, 1986

Talks on testing at conferences sponsored by ETS and by the College Board
Several talks at the AERA annual meeting

July, 1986

A talk on prediction/postdiction at the ONR Contractors meeting in Champaign, IL
Lectures and workshops on thinking at the Carkhuff Institute (Amherst), the College of St. Thomas (St. Paul), the Rotary Club (Hamden, CT), Confratute (Storrs, CT), Fairfield University, an ASCD institute in Williamsburgh, Northwestern University, and Woodbury (CT).

August, 1986

A colloquium on theory of intellectual styles at Duke University
An invited address and 2 panel discussions at APA in Washington
A lecture on intelligence at the University of Leuven (Belgium)
An invited address at the annual meeting of the German Psychological Society in Heidelberg

October, 1986

A workshop on critical thinking in Newtown (CT) High School
A workshop on testing intelligence to the Connecticut School Psychologists Association
A talk on testing for college asmissions at the College Board annual conference
A lecture on my theory of intelligence in Wilkesboro, NC, the site for the pilot Multidimensional Ability Test
A lecture on thinking in the West Irondequoit School District (upper New York State).

November, 1986

A keynote address to the New York State Reading Association

A lecture on thinking skills to textbook writers and editors at Harcourt Brace Jovanovich International headquarters in Orlando

A talk on novelty and intelligence at the Psychonomic Society meetings

A major address and panel discussion at a Canadian Conference on thinking skills in Toronto

December, 1986

An all-day workshop on thinking at the University of California, Irving

January, 1987

An invited address on thinking skills at the annual North Carolina ASCD meeting

Talks to a series of educational consortia in New York State through N.Y. Central BOCES.

February, 1987

Colloquia at Quinnipiac College (New Haven), the Education School of the University of Pennsylvania, and the University of Washington Psychology Department

A talk to school district officials in Seattle through a program arranged by Antioch University

A talk at an ETS meeting on graduate student success

March, 1987

Colloquia at the University of Saskatchewan and of Regina

Workshops at the University of Puerto Rico in Rio Piedras and in Cayley

A talk at a conference on schooling and intelligence at the University of Michigan

A talk at a meeting on personnel and instruction in San Antonio

Workshops on thinking skills in Trumbull (CT), Washtinaw County (MI), and Mamaroneck, NY

The keynote address and workshop at a conference on giftedness at Arlington, TX

A talk at the ONR Contractors' meeting

April, 1987

Colloquia on my research on intelligence at the University of Texas, San Antonio, the University of Maryland, and the University of Pittsburgh

May, 1987
- A colloquium at the City University of New York
- Talks at a New Jersey State Conference on Thinking Skills, the New York ASCD annual conference, and the University of Minnesota Round Table

June, 1987
- A talk at the annual ONR Contractors' meeting
- A workshop in the Fairfield (CT) schools

July, 1987
- An invited address on my views of intelligence and its development to the International Society for the Study of Behavior Development in Japan
- Two lectures on the nature of intelligence at the University of Hawaii

August, 1987
- An invited address on the nature of giftedness at the World Gifted Conference in Salt Lake City
- A three-day workshop on thinking skills sponsored by AERA in San Francisco
- An invited address on intelligence testing and a symposium on aging at APA annual conference in New York

September, 1987
- A colloquium at the State University of New York in Buffalo

October, 1987
- A keynote address at the Florida State Gifted Association annual meeting
- A series of lectures to educators in the St. Louis Special School District
- Workshops in the New Fairfield and South Windsor (CT) School Districts

November, 1987
- A talk at the annual meeting of the Psychonomic Society
- A colloquium at Carnegie-Mellon University
- The keynote address at the Tennessee and Illinois State Gifted Associations annual meetings
- A workshop to high-level educators at Research for Better Schools in Philadelphia

December, 1987
- A colloquium at the University of Cincinnati

February, 1988
- A colloquium at the University of Toledo
- A colloquium at Florida State University
- An invited address on testing to the National Association of Independent Schools in NYC
- A university-wide lecture at Bowdoin College

March, 1988
- An address to a consortium of educators from the Chicago region
- A talk to teachers at the Bronxville School
- An address to scholars at the University of Texas, Austin, in a conference on new directions in testing organized by Earl Hunt and Don Foss

April, 1988
- A colloquium at the University of Rochester
- A colloquium at the University of Wisconsin, River Falls
- A talk at a conference at Calgary
- A talk at Columbia High School in Maplewood, NJ

May, 1988
- A talk at a high school in Illinois

June, 1988
- A talk to testing personnel of the U.S. Employment Service

July, 1988
- Colloquia and talks at an International Conference on Individual Differences in Cognition in Ramat Gan, Isreal and at University of Haifa
- Talks at the National Institute of Testing and Evaluation in Jerusalem and the Weizman Institute of Science
- A week-long institute on thinking skills at Yale
- A talk to Confratute (an institute for teachers of the gifted)

August, 1988
- Two talks at the American Psychological Association meetings in Atlanta
- A talk at the International Conference on Intelligence sponsored by the Australian government
- A talk at the International Conference on Thinking Skills in Edmonton, Alberta